MOUNT SINAI
GUSTAVE L.
AND
JANET W.
LEVY
LIBRARY

Digitized by the Internet Archive
in 2015

https://archive.org/details/mountsinaijourna5819moun

# THE MOUNT SINAI JOURNAL OF MEDICINE

OF THE CITY UNIVERSITY OF NEW YORK

OLUME 58 NUMBER 1 JANUARY 1991

# The Art and Science of Consultation in Medicine

Paul L. Gilbert, guest editor

From the Division of General Medicine,
Department of Medicine, Mount Sinai School of Medicine

Veterans Administration Medical Center
Bronx, New York

BETH
ISRAEL
MEDICAL
CENTER

**Where there's smoke...there may be bronchitis**

"Recent research has delineated early, more subtle changes in lung and immune functions. These alterations directly predispose smokers to respiratory tract infection."

*Am Fam Phys* 1987;36:133-140

Ceclor® Pulvules® 250 mg
cefaclor

*Established therapy for today's patients*

For respiratory tract infections due to susceptible strains of indicated organisms

**Brief Summary.**
**Consult the package literature for prescribing information.**
**Indication:** Lower respiratory infections, including pneumonia, caused by *Streptococcus pneumoniae*, *Haemophilus influenzae*, and *Streptococcus pyogenes* (group A β-hemolytic streptococci).
**Contraindication:** Known allergy to cephalosporins.
**Warnings:** CECLOR SHOULD BE ADMINISTERED CAUTIOUSLY TO PENICILLIN-SENSITIVE PATIENTS. PENICILLINS AND CEPHALOSPORINS SHOW PARTIAL CROSS-ALLERGENICITY. POSSIBLE REACTIONS INCLUDE ANAPHYLAXIS.

Administer cautiously to allergic patients.

Pseudomembranous colitis has been reported with virtually all broad-spectrum antibiotics. It must be considered in differential diagnosis of antibiotic-associated diarrhea. Colon flora is altered by broad-spectrum antibiotic treatment, possibly resulting in antibiotic-associated colitis.

**Precautions:**
- Discontinue Ceclor in the event of allergic reactions to it.
- Prolonged use may result in overgrowth of nonsusceptible organisms.
- Positive direct Coombs' tests have been reported during treatment with cephalosporins.
- Ceclor should be administered with caution in the presence of markedly impaired renal function. Although dosage adjustments in moderate to severe renal impairment are usually not required, careful clinical observation and laboratory studies should be made.
- Broad-spectrum antibiotics should be prescribed with caution in individuals with a history of gastrointestinal disease, particularly colitis.
- Safety and effectiveness have not been determined in pregnancy, lactation, and infants less than one month old. Ceclor penetrates mother's milk. Exercise caution in prescribing for these patients.

**Adverse Reactions:** (percentage of patients)

Therapy-related adverse reactions are uncommon. Those reported include:
- Hypersensitivity reactions have been reported in about 1.5% of patients and include morbilliform eruptions (1 in 100). Pruritus, urticaria, and positive Coombs' tests each occur in less than 1 in 200 patients. Cases of **serum-sickness-like** reactions have been reported with the use of Ceclor. These are characterized by findings of erythema multiforme, rashes, and other skin manifestations accompanied by arthritis/arthralgia, with or without fever, and differ from classic serum sickness in that there is infrequently associated lymphadenopathy and proteinuria, no circulating immune complexes, and no evidence to date of sequelae of the reaction. While further investigation is ongoing, **serum-sickness-like** reactions appear to be due to hypersensitivity and more often occur during or following a second (or subsequent) course of therapy with Ceclor. Such reactions have been reported more frequently in children than in adults with an overall occurrence ranging from 1 in 200 (0.5%) in one focused trial to 2 in 8,346 (0.024%) in overall clinical trials (with an incidence in children in clinical trials of 0.055%) to 1 in 38,000 (0.003%) in spontaneous event reports. Signs and symptoms usually occur a few days after initiation of therapy and subside within a few days after cessation of therapy; occasionally these reactions have resulted in hospitalization, usually of short duration (median hospitalization = two to three days, based on postmarketing surveillance studies). In those requiring hospitalization, the symptoms have ranged from mild to severe at the time of admission with more of the severe reactions occurring in children. Antihistamines and glucocorticoids appear to enhance resolution of the signs and symptoms. No serious sequelae have been reported.
- Stevens-Johnson syndrome, toxic epidermal necrolysis, and anaphylaxis have been reported rarely. Anaphylaxis may be more common in patients with a history of penicillin allergy.
- Gastrointestinal (mostly diarrhea): 2.5%
- Symptoms of pseudomembranous colitis may appear either during or after antibiotic treatment.
- As with some penicillins and some other cephalosporins, transient hepatitis and cholestatic jaundice have been reported rarely.
- Rarely, reversible hyperactivity, nervousness, insomnia, confusion, hypertonia, dizziness, and somnolence have been reported.
- Other: eosinophilia, 2%; genital pruritus or vaginitis, less than 1% and, rarely, thrombocytopenia and reversible interstitial nephritis.

Abnormalities in laboratory results of uncertain etiology.
- Slight elevations in hepatic enzymes.
- Transient lymphocytosis, leukopenia, and, rarely, hemolytic anemia and reversible neutropenia.
- Rare reports of increased prothrombin time with or without clinical bleeding in patients receiving Ceclor and Coumadin concomitantly.
- Abnormal urinalysis; elevations in BUN or serum creatinine.
- Positive direct Coombs' test.
- False-positive tests for urinary glucose with Benedict's or Fehling's solution and Clinitest® tablets but not with Tes-Tape® (glucose enzymatic test strip, Lilly).

PA 8791 AMP [021490LRI]

*Additional information available to the profession on request from Eli Lilly and Company, Indianapolis, Indiana 46285.*

Lilly

Eli Lilly Industries, Inc
Carolina, Puerto Rico 00630
A Subsidiary of Eli Lilly and Company
Indianapolis, Indiana 46285

CR-0525-B-049333 © 1990, ELI LILLY AND COMPANY

]

HE
OUNT SINAI JOURNAL
F MEDICINE

VOLUME 58
NUMBER 1
JANUARY 1991

# CONTENTS

## THE ART AND SCIENCE OF CONSULTATION IN MEDICINE

**Editorial** *Sherman Kupfer* . . . . . 1

**Introduction: In Omnia Paratus——Ready for All Things** *Paul L. Gilbert* . . . . . 2

**The Internist as Preoperative Consultant: Risk Assessment and Management** *Paul L. Gilbert* . . . . . 3

**Anesthesia: What the Internist Needs to Know** *Andrew B. Leibowitz and Jeffrey H. Silverstein* . . . . . 9

## PREOPERATIVE PREVENTION

**Antimicrobial Therapy** *Eric Neibart* . . . . . 19

**Prevention of Deep Venous Thrombosis and Pulmonary Embolism in Surgical Patients** *Jonathan L. Halperin* . . . . . 28

## PREOPERATIVE EVALUATION

**Pulmonary Status** *Jonathan T. Klein and Mark J. Rosen* . . . . . 37

**Hematologic Status** *Cesare Calandri and Jacob H. Rand* . . . . . 41

**The Patient with Cardiac Disease** *Herschel J. Sklaroff* . . . . . 48

**The Patient with Endocrine Disease** *Paul L. Gilbert* . . . . . 58

**The Patient with Chronic Renal Disease** *Paul L. Gilbert and Richard Stein* . . . . . 69

**The Patient with Liver Disease** *Mitchell Conn* . . . . . 75

**The Pregnant Patient: Need for Surgery Unrelated to Pregnancy** *Sheldon H. Cherry* . . . . . 81

**The Elderly Psychiatric Patient** *Paul S. Aisen* . . . . . 85

## BRIEF COMMUNICATION

**Physician-Nurse Collegiality in the Medical School Curriculum: Exploratory Workshop and Student Questionnaire** *Patricia Lynch Anvaripour, Linbania Jacobson, Joyce Schweiger, and Gail Kuhn Weissman* . . . . . 91

**Index to Advertisers**

**Eli Lilly and Company** **cover 2**

# THE MOUNT SINAI JOURNAL OF MEDICINE

The Mount Sinai Journal of Medicine is published by The Mount Sinai Medical Center of New York and has the following affiliates: Beth Israel Medical Center, New York; Bronx Veterans Administration Medical Center, New York; and Elmhurst Hospital Center, New York.

**Editor-in-Chief**

Sherman Kupfer, M.D.

**Editors Emeriti**

David A. Dreiling, M.D.    Lester R. Tuchman, M.D.

**Associate Editors**

Harriet S. Gilbert, M.D.    Julius Wolf, M.D.

**Managing Editor**

Claire Sotnick

**Assistant Editors**

David H. Bechhofer, Ph.D.
Constanin A. Bona, M.D., Ph.D.
Edward J. Bottone, Ph.D.
Jurgen Brosius, Ph.D.
Steven T. Brower, M.D.
Lewis Burrows, M.D.
Joseph S. Eisenman, Ph.D.
Richard A. Frieden, M.D.
Steven Fruchtman, M.D.
Paul L. Gilbert, M.D.
James H. Godbold, Ph.D.
Richard S. Haber, M.D.
Noam Harpaz, M.D.
Dennis P. Healy, Ph.D.
Barry W. Jaffin, M.D.
Andrew S. Kaplan, D.D.S.
Samuel Kenan, M.D.
Suzanne Carter Kramer, M.Sc.
Thomas A. M. Kramer, M.D.
Mark G. Lebwohl, M.D.
Kenneth Lieberman, M.D.
Charles Lockwood, M.D.
Bernard Mehl, D.P.S.
Myron Miller, M.D.
Edward Raab, M.D.
Allan Reed, M.D.
Allan E. Rubenstein, M.D.
David B. Sachar, M.D.
Henry Sacks, M.D.
Robert Safirstein, M.D.
Ira Sanders, M.D.
Martin H. Savitz, M.D.
Clyde B. Schechter, M.D.
Phyllis Shaw, Ph.D.
George Silvay, M.D.
Barry D. Stimmel, M.D.
Nelson Stone, M.D.
Max Sung, M.D.
Rein Tideiksaar, Ph.D.
Richard P. Wedeen, M.D.
Michael F. Wesson, M.D.

**Editorial Board**

Barry Freedman, M.B.A.
Richard Gorlin, M.D.
Nathan Kase, M.D.
Panayotis G. Katsoyannis, Ph.D.
Charles K. McSherry, M.D.
Jack G. Rabinowitz, M.D.
John W. Rowe, M.D.
Alan L. Schiller, M.D.
Alan L. Silver, M.D.
Alvin S. Teirstein, M.D.

*The Mount Sinai Journal of Medicine* (ISSN No. 0027-2507; USPS 284-860) is published 6 times a year in January, March, May, September, October, and November in one indexed volume by the Committee on Medical Education and Publications (Owner), The Mount Sinai Hospital, Box 1094, One Gustave L. Levy Place, New York, NY 10029-6574. Subscription price: individuals, U.S. and Canada, $60 per year; all libraries, $65; individuals outside the U.S. and Canada, $65. Single copies, $15. New subscriptions begin with the first issue of the calendar year. Send notice of change of address 60 days before the change is effective. Second-class postage paid at New York, NY and at additional mailing offices. POSTMASTER: Send address changes to The Mount Sinai Journal of Medicine, Box 1094, One Gustave L. Levy Place, New York, NY 10029-6574. Copyright 1991 The Mount Sinai Hospital.

Statement of Ownership, Management, and Circulation filed with the U.S. Postal Service. Average no. copies each issue during preceding 12 months: 10a. net press run, 3071; 10b. 1. single copy sales, 20; 10b. 2. mail subscriptions, 2695; 10c. total paid or requested circulation, 2715; 10d. nonpaid distribution, 75; 10e. total distribution, 2790; 10f. 1. copies not distributed, 281; 10f. 2. return from news agents, none; 10g. total number of copies, 3071. Actual no. of copies of issue published nearest to filing date: 10a. net press run, 2970; 10b. 1. single copy sales, 50; 10b. 2. mail subscriptions, 2735; 10c. total paid or requested circulation, 2785; 10d. nonpaid distribution, 80; 10e. total distribution, 2865; 10f. 1. copies not distributed, 105; 10f. 2. return from news agents, none; 10g. total number of copies, 2970.

## Editorial

Sherman Kupfer, MD

# The Art and Science of Consultation In Medicine

WITH this issue the *Journal* hopes to initiate a series entitled "The Art and Science of Medical Consultation." The current issue does not cover consultation by internists in all specialties, nor was it intended to do so. Also, it is not encyclopedic. In fact, the basis of this issue derives, as Paul Gilbert says in his introduction to the issue, from a series of presentations made to The Mount Sinai Hospital housestaff in the Department of Internal Medicine during their time assigned to the Consultation Service of the Division of General Medicine. During the past several years under his leadership, the preceptors have been encouraged to include in their presentations to the housestaff some written material. Now this material has been collated and updated. We hope that this initial effort will encourage many consultants—at the Mount Sinai Medical Center and elsewhere—to prepare scholarly writings on their experiences, so that the art as well as the science can be transmitted through the pages of the *Journal* to its readers.

The Division of General Medicine and its assigned housestaff have the responsibility of providing medical consultation on request to the other services, particularly the surgical services. Although there is a high degree of unanimity among the consultants in the articles published here, there are important differences in both approach and style. Many of these differences are based on the perceived lack of conclusive data. In this situation, the consultant must make educated guesses as to what might be the best treatment approach for the specific patient for whom the consultation has been called. These conclusions must then be presented to the referring physician so as to guide in the selection of available alternatives. The manner in which this information is communicated, the promptness of the communication, and the nature of the interaction between the physicians are all crucial to the effectiveness of the consultation.

A second critical aspect of the art of consultation is the interaction between the consultant and the patient. Perhaps this is less of a problem on the in-hospital service. Too often in this situation, the patient is unaware of the reason why a consultant has been called. Perhaps too little attention has been paid to this aspect, since the consultant may think that his or her sole responsibility is to transmit highly technical information to the referring physician. Little or no information may be given directly to the patient. It is not surprising that the patient may not even recall the visit.

Another problem arises when the patient is out of the hospital and undertakes to arrange for a "second opinion" without informing the original physician. This situation will become apparent to the consultant who asks the question, "Who sent you to me?" What kind of dilemma does this pose for the consultant in respect to the kind of interaction which must ultimately take place between the patient and the consultant and the consultant and the original physician? This facet too has not been addressed in these articles.

There is no question that the major focus of the group of articles presented in this issue is the quest for a complete cure—which would certainly rank as a successful outcome to any proposed therapy. Surely, that is what is uppermost in the mind of the physician who requests a consultation and the patient. Unfortunately, it cannot be guaranteed, since in many circumstances the consultant can offer little in the way of an optimistic treatment plan. Instead, the consultant is left to offer a plan of palliation, relief, and comfort. Few would deny that how this information is presented to the referring physician and to the patient and his or her family is truly an art.

It is possible that not every consultant specialist has to address such an array of concerns, but each specialty may have problems unique to it and the art to deal with them.

## *Introduction*

Paul L. Gilbert, MD

# In Omnia Paratus—Ready for All Things

IN A TERTIARY care institution such as the Mount Sinai Medical Center, in which the prevailing wisdom is that our residents are conferenced to death and their attendance even at scheduled conferences is only as regular as their heavy clinical responsibilities permit, the phenomenon of residents requesting an added lecture course in Consultation Medicine during their lunch hour was met with some skepticism by the faculty. In response to continued prodding by our chief residents, such a series of lectures, comprising the contents of most of this issue, was arranged and, surprisingly, was well attended. The course was not designed to recapitulate the many excellent texts on the subject, nor was the content of the lectures directed at the subspecialist. It was designed to be a practical approach for the generalist, and the editors are aware that *non omnia possusmus amnes*—we cannot do all things.

Readers will note variations in opinion on, for example, the perioperative use of insulin, steroids, anticoagulants, antibiotics. No attempt has been made to unify, reconcile, or institutionalize approaches to the various problems on which opinion varies. Definitive solid data are often lacking, and many of these questions lend themselves to various approaches.

Some explanation for a rekindling of interest in general medical curriculum may be found in several factors:

1. Residents observe that the lack of orchestration of the input of multiple subspecialists consulting on surgical patients is often cost ineffective and sometimes leads to conflicting orders and confusing or poorly prioritized recommendations for further evaluation.
2. The housestaff perceives a constriction in their exposure to hospitalized patients as beds in teaching hospitals in metropolitan areas are increasingly filled with the hyperacutely ill, those with end-stage disease, and those admitted for surgical procedures. Many patients with bread-and-butter diseases are no longer being admitted. Critical decision-making and studies to solve diagnostic problems are increasingly being carried out in the ambulatory setting. More and more patients are being hospitalized for day-of-admission surgery, thus creating a demand for rapid and effective preoperative evaluation without the luxury of time for involved diagnostic study.
3. Local, state, and federal support of postgraduate education continues to dwindle, while demands for improved access to and quality of service within shrunken budgets increase. As educators struggle with the problems of funding such education and of adjusting programs to keep them relevant to the needs of residents who may opt for either academic or clinical careers and at the same time to maintain the programs within the mandates of the board examinations, our housestaff is dealing with its own set of problems. The climate created by the alphabet of acronyms of medicine of the 1990s (DRGs, HMOs, PROs, and more to come) and the decreased funding of education has made residents' future roles as subspecialists unclear. Residents recognize as likely that, except for those among their ranks choosing highly procedure oriented specialties or specialties in which bench research is predominant, the majority of them going into practice will do so in groups, where the economics of new reimbursement systems will dictate that they do general medicine as well as their subspecialty. They correctly perceive improved general consultation skills as an added need.

---

The author is Chief, Division of General Medicine, Department of Medicine, Mount Sinai School of Medicine, New York, NY.

# The Internist as Preoperative Consultant: Risk Assessment and Management

PAUL L. GILBERT, M.D.

---

IT IS the function of the medical consultant on the surgical services to identify underlying medical problems, to categorize and estimate the operative risk factors, to maximize the patient's medical condition preoperatively, and to predict and assist in avoiding and managing intraoperative and postoperative problems.

### Non nova sed nove—Not new but anew

Medical consultation on the surgical services is a refocusing of the broad knowledge base of internal medicine onto a specific problem—preoperative risk evaluation. There is a modest but growing literature on the subject. An effective consultation should incorporate these data. A consultation must be formulated in the context of good clinical judgment based on understanding of the pathophysiology of the underlying disease, the pharmacology of the medications to be used, and their potential interactions with the physiologic changes attendant on surgery and anesthesia.

## Risk Assessment

Over 20,000,000 Americans undergo anesthesia and surgery each year. On one hand, over 2% of them die during the same hospital admission as the surgery. On the other, over half of them fit into American Society of Anesthesiologists Class 1, that is, they have no disease other than the surgical pathology (Table). The overall surgical mortality in this group is 0.06%, depending on (a) the location of the surgery (highest in intracranial, intrathoracic, and intraabdominal procedures), (b) whether elective or emergency (mortality increases 3–4 times in emergency surgery), and (c) the age of the patient. The higher mortality in patients over age 70 is likely a reflection of the high incidence of emergency surgery and underlying degenerative and neoplastic disease in this group.

Many of the studies quoted in this area are several years old. Conclusions drawn from them should be modified by the impact of newer medications and therapies, improved surgical techniques, and our ability to monitor invasively and intervene more effectively. For instance, the concept that reinfarction rates are 27%–37% in patients operated on within 3 months of a myocardial infarction, 11%–16% in those being operated on within 4–6 months, and 4%–5% when the operation takes place after 6 months has been challenged by Rao, who, using invasive monitoring to dictate early and aggressive therapeutic response to changes in baseline values, noted only a 5.7% reinfarction rate if the operation was performed within 3 months of a myocardial infarction, and 2.3% if it was at 4–6 months. The application of newer methods of evaluating the integrity of remaining myocardium, ventricular function, and the extent and location of previous damage may, in the future, lend better insight into the subgroups at greater or lesser risk. The constancy of surgical mortality in the aged may be more a function of our increasing willingness to perform more major surgery on sicker elderly patients than to any necessarily fixed percentage of mortality following surgery in this age group.

No extensive discussion is offered here of the preoperative evaluation of the healthy young patient. Synthesizing the current literature, a strong statistical argument can be made for an extreme position: no added evaluation by an internist is necessary in a young patient who feels

---

From the Division of General Medicine, Department of Medicine, The Mount Sinai Medical Center, New York, NY 10029. Address reprint requests to the author, who is Chief, Division of General Medicine, at the Department of Medicine, Box 1088, Mount Sinai School of Medicine, One Gustave L. Levy Place, New York, NY 10029.

TABLE
*Physical Status Classification*

| Physical status | Description of patient | 7-day postop mortality (%) |
|---|---|---|
| 1 | Normal healthy | 0.06 |
| 2 | Has mild systemic disease | 0.4 |
| 3 | Has physical disease limiting activity but not incapacitating | 4.3 |
| 4 | Has incapacitating disease that threatens life | 23.4 |
| 5 | Moribund, not expected to survive 24 hr even with operation | 50.7 |

Adapted from American Society of Anesthesiologists Physical Status Classification and Marx's postoperative mortality studies.

well, looks healthy from the foot of the bed, answers "no" to a series of 13 questions related to respiratory, cardiac, and hemostatic function, has had no significant past illnesses, takes no medications, is not pregnant, and has not had anesthesia in the prior three months or any personal or family history of problems with anesthesia. Aside from urinalysis, hemoglobin or hematocrit, blood urea nitrogen, and pregnancy test (in appropriate patients), further preoperative laboratory study is not cost effective. If the answers to such a questionnaire were all negative in 200 consecutive patients and an electrocardiogram (ECG) were taken on each, one ECG abnormality would be found which might cause a delay in surgery. The specificity and predictability of the ECG in older patients makes it unreliable in predicting intra- and post-operative complications. Much the same ratio of abnormalities uncovered can be demonstrated for routine preoperative chest X-rays and for hemostatic testing.

Despite such studies, many clinicians have serious problems with such a minimalist approach to preoperative studies and believe that the answer lies between the two extreme points of view. The exponential cost of running a SMAC 12 is, at most, several dollars. The greatest expenditure is not the group of tests, but the almost 50% chance of one of the tests showing a spurious abnormality, further study of which could be costly. Here, clinical judgment and prioritization of diagnostic studies are primary. Although much maligned as being cost inefficient, screening panels have an important place in medical practice. Improved early case finding in hyperparathyroidism is but one example.

The ever-increasing financial pressures placed on medicine by diagnosis-related groups (DRGs) have too often distorted the meaning of *cost-effective* to the equivalent of *cheaper*. What would be the cost in human suffering, further medical care, malpractice suits, and peer review organization (PRO) proceedings if a preoperative survey missed important data in a patient who was a poor historian, if laboratory evidence of severe hypokalemia induced by surreptitious use of laxatives and diuretics was not picked up, if an unrecognized mild deficit of Von Willebrand factor became magnified at surgery and resulted in a major intra- or postoperative problem with hemostasis, if there was postoperative bacterial seeding on an unrecognized mitral valve prolapse or other asymptomatic diseased valve? The seasoned clinician recognizes the limitations of his or her experience and the validity of analyses of large numbers of cases. However, the intellectual and emotional impact of one's individual experiences cannot be voided by statistics, and heuristic thinking often dominates—the "once burned, twice cautious" aphorism often takes precedence.

## Effects of Anesthesia

Although the choice of anesthetic agents is properly in the realm of the anesthesiologist, it is important for the consulting internist to understand the physiologic alterations accompanying anesthesia as she or he recommends any intervention during this period.

We would rather not subject many high-risk patients to the stresses of anesthesia or surgery. Internists often lean toward regional anesthesia for such patients. It may be that under certain circumstances (perineal, lower extremity, rectal, and perhaps lower abdominal surgery) such an approach may add a margin of safety. However, spinal anesthesia results in vasodilation (increasing the propensity to hypotension, particularly in the presence of volume depletion), it alters cardiac pre- and afterload, does not protect against aspiration, can interfere with respirations (high spinal), and, since these agents are metabolized by the liver, may have prolonged effects in patients with dysfunction of this organ. It is contraindicated in many neurologic diseases, in the presence of bleeding diathesis, sepsis, and local infection of the back. Neuroleptoanesthesia, which offers a balanced anesthesia with good cardiovascular stability, may be a preferable alternative for major procedures in high-risk patients.

All general anesthetics depress myocardial function. Cardiac output and stroke volume indices decrease in proportion to the dose given. Catecholamine secretion is increased. Baro- and chemo-receptor function is depressed. Peripheral vasodilation is produced and results in relative hypovolemia. The potential for lability of blood pressure, hypotension, tachycardia, and arrhythmias is created.

Anesthesia results in respiratory depression. There is a blunting of the central nervous system (CNS) responses to hypercarbia and hypoxia. Pulmonary compliance and lung volumes, particularly tidal volume and functional residual capacity, decrease. Via several mechanisms, ventilation-perfusion mismatches occur and can result in increased a-A $O_2$ gradients and hypoxia. Positive pressure ventilation decreases venous return and, especially in hypovolemic states, can contribute to hypotension. The cough reflex and mucociliary escalator mechanisms are depressed, and viscosity of secretions increases.

Renal blood flow decreases up to 50%; glomerular filtration rate, renal vascular resistance, and filtration fraction may decrease; and normal autoregulation is disrupted by inhalation agents. Liberation of vasopressor substances and ADH, and activation of the renin-angiotensin system with aldosterone effects, all have an impact on these hemodynamic alterations. Most anesthetic drugs are not eliminated in an active form by the kidney, having first been metabolized in the liver. However, the highly ionized, nondepolarizing muscle relaxants (gallamine, d-tubocurarine, metocurine), neostigmine, and the ganglionic blockers are eliminated unchanged by the kidneys and must be used with caution in patients with chronic renal disease. Cholinesterase levels are low in uremic patients (even lower in the dialyzed patient), and as a result, the effects of the neuromuscular blockading agents may persist longer into the postoperative period than anticipated, producing further problems with hypoventilation.

All inhalation anesthetics decrease hepatic blood flow and hepatic oxygen uptake. Mechanical ventilation (by altering $CO_2$ levels), decreased cardiac output, and splanchnic reflexes (secondary to surgical manipulation) enhance this effect. Almost all potent anesthetics are biotransformed by the liver. Anatomically, the closer the surgery to the liver, the more likely are minor degrees of postoperative elevation of serum transaminase, independent of the anesthesia utilized. Drugs with properties that induce liver enzymes (alcohol, phenobarbital) may increase the anesthetic dose required; those metabolized by the liver or cleared unchanged by the kidney will have a prolonged half-life in patients with diseases of these organs. The hypoproteinemia of severe liver disease results in decreased protein binding of the anesthetic, further accentuating its effect in patients so afflicted.

In general, the decrease in muscle tone induced by neuromuscular blockers can result in blood pooling and hypotension. Respiratory muscle function is paralyzed until these drugs are cleared from the system. The depolarizing blockers are metabolized by pseudocholinesterase, and their effects remain until the body produces a sufficiency of this enzyme. The nondepolarizing agents are reversible by cholinesterase inhibitors such as neostigmine; their effects are potentiated by the ganglionic blockers. The calcium channel blockers potentiate the effects of all of the neuromuscular blocking agents, and the aminoglycosides compete with all of them.

General anesthesia results in some deterioration of glucose metabolism. Glycogenolysis, insulin-reversible inhibition of glucose transport, and decreased insulin release all contribute. ACTH secretion increases with induction, followed in 30 min by a rise in cortisol levels, which then further increase with intubation and surgery. There is a rise in aldosterone, ADH, and catecholamine production in association with general anesthesia.

### Managing Medications Preoperatively

Few hard and fast rules govern beginning medication and stopping it prior to surgery. Reliable clinical studies are scarce, the literature is not rich, and the efficacy of continuation or addition of certain medications in the perioperative period remains moot. The integrated input of surgeon, anesthesiologist, and internist is essential. Each patient must be managed individually, taking into consideration the nature of the underlying medical condition(s), the type and urgency of the surgery, the pharmacologic effects of the medications in question, their possible interactions with agents used during surgery, and their possible effects on the altered physiology of the perioperative period.

The individual contributors to this issue have been asked to deal with the starting and stopping of pertinent medications in their discussions and I do not dwell heavily on them here except in generalities.

Moderate hypertension does not appear to be an independent operative risk factor if there is no target organ involvement. Patients being treated by antihypertensive drugs are more vulnerable to marked blood pressure fluctuations during surgery, particularly if blood volume is depleted. Diuretics predispose the patient to hypokalemia and volume depletion. Congestive heart failure is a major preoperative risk factor, and the internist dealing with a patient in failure who requires urgent surgery must walk the razor's edge in accommodating the needs for and risks of diuretic therapy. The sudden discontinuation of beta blockers and clonidine can result in life-threatening discontinuation syndromes. Most of the other antihypertensives can potentiate several of the undesirable effects of anesthetics and surgery and are best discontinued the night before operation. Intraoperative hypertension is manageable with parenteral hydralazine, methyldopa, beta blockers, or, if severe or unresponsive, by nitroglycerine, nitroprusside, or phentolamine.

Nitrates should be continued as medically indicated. The sublingual route may be used pre- and post-operatively; the transdermal or intravenous (IV) routes (with proper monitoring), intraoperatively.

Patients with asthma and chronic obstructive pulmonary disease should be rendered wheeze-free preoperatively. Whenever possible, bronchodilators should be discontinued 12 hr prior to surgery (to avoid arrhythmogenic and hypotensive effects). Many pulmonologists recommend preoperative stress doses of cortisol. Maximal obviation of the risks of smoking probably requires discontinuation 3–4 weeks prior to surgery.

Digitalis should be continued in those patients who require the drug for rate control of arrhythmias, keeping in mind that the altered physiology during surgery predisposes to digitalis toxicity. Verapamil and Inderal are effective in the management of intraoperative supraventricular tachycardias (SVTs). The presence of more than five ventricular premature contractions (VPCs)/minute represents a risk factor for intra- and post-operative morbidity only when they are manifestations of underlying heart disease; otherwise they are relatively innocuous. Paradoxically, the major risk they impose is not so much due to the development of serious arrhythmias but of other cardiovascular complications (myocardial and pulmonary infarctions, congestive heart failure, and so on). Patients who have experienced "sudden death," symptomatic ventricular arrhythmias, and perhaps those with frequent, complex, multifocal ventricular extrasystoles may be given a dose of the antiarrhythmic with a sip of water early on the morning of surgery, and/or they may receive prophylactic intravenous procainamide during surgery. As a group, the ventricular antiarrhythmics act as myocardial depressants and can cause hypotension. They should be discontinued in other patients the night prior to operation and resumed when oral intake is permitted.

Aspirin irreversibly acetylates platelet cyclooxygenase, thus inhibiting aggregation for the life of the platelet. It should be discontinued 7–10 days preoperatively. The other nonsteroidals affect platelet function transiently and need be discontinued only 2–3 days prior to surgery. Parenteral carbenicillin, ticarcillin, and moxalactam can cause platelet dysfunction and bleeding.

Well-controlled hyperthyroidism and hypothyroidism present no exponential operative risks and maintenance therapy can be discontinued for several days without ill effect. Uncontrolled or poorly controlled hyperthyroidism presents significant operative risk and, if surgery cannot be deferred, requires aggressive preoperative therapy. Mild to moderate untreated hypothyroidism does not always require preoperative replacement, but severe hypothyroidism usually requires some intravenous $T_4$ and cortisol preoperatively.

Chlorpropamide should be discontinued 2–3 days preoperatively because of its prolonged potential for hypoglycemia. The other oral hypoglycemics may be discontinued the night prior to surgery. Perioperative use of insulin is discussed fully in the article on evaluating the patient with endocrine disease.

Corticosteroids in suppressive doses given for 5–7 days or more can inhibit the response of the pituitary-adrenal axis for a year. Stress doses of cortisol are indicated if any question exists.

Tranquilizers in large doses may act as cardiac depressants and potentiate those effects of anesthesia. They should be discontinued 2–3 days preoperatively. Lithium can prolong the effect of muscle relaxants (via its effect on membrane potentials), can cause sodium retention, induce nephrogenic diabetes insipidus, and result in hypothyroidism (by inhibiting hormonal release from the thyroid gland). It is best discontinued 5–7 days prior to operation.

The monoamine oxidase inhibitors behave erratically with multiple side effects and drug interactions. They are best stopped 1–2 weeks prior to surgery. Their pharmacologic effects increase the potential for sympathetic hyperactivity and reactivity. Hypertension, hypotension, convulsions, and coma may occur. They may potentiate the effects of the oral hypoglycemics, and interactions with L-dopa have been reported.

The tricyclic antidepressants increase the CNS response to depressants and peripheral sensitivity to sympathetic agonists; have significant anticholinergic activity; are arrhythmogenic

(A-V conduction defects); can cause orthostatic hypotension (especially in the elderly); and may act as cardiac depressants potentiating the effects of lidocaine and quinidine. Hyperthermia has been observed to follow the administration of the tricyclic antidepressants. The use of these agents should also be curtailed 1–2 weeks before scheduled surgery.

Phenytoin levels should be checked preoperatively and maintained at therapeutic levels intra- and post-operatively. Any major motor seizure may result in aspiration pneumonia, wound dehiscence, or disruption of implanted prostheses or reduced fractures. Since the intramuscular preparation is absorbed irregularly, the intravenous route is used until the patient is able to resume oral feedings. One or two doses of phenobarbital can be missed without ill effect. In situations in which a patient is taking another anticonvulsant medication that is not available in parenteral form, phenytoin and phenobarbital may be substituted if the half-life of the drug in question is insufficient to protect against seizures until oral intake is possible. All but life-saving surgery should be deferred for 2 weeks after a cerebrovascular accident, when cerebral autoregulation has normalized.

The internist acting as medical consultant frequently must make preoperative evaluations of patients who are taking anticoagulant medications. Since long-term anticoagulation is usually undertaken as prophylaxis for potentially life-threatening thromboembolic disease and adequate hemostasis is fundamental to all surgery, the internist finds himself or herself in a therapeutic dilemma. The literature is somewhat confusing and in many situations does not offer clear-cut answers. Good clinical judgment, taking into consideration the level of risk of thromboembolic events from underlying conditions and the type of surgery, will mandate the approach.

Clearly, the tolerance for any oozing is extremely limited in ophthalmologic, plastic, and neurologic procedures; such patients are generally operated on fully corrected. When necessary, many minor operations, such as dental extractions, soft-tissue biopsy, and arteriography, may be performed under full anticoagulation. Certain procedures, particularly orthopedic surgery of the lower extremities, are accompanied by a sufficiently high risk of thromboembolic complications as to warrant considering maintenance of a low level of anticoagulation throughout the entire perioperative period. If the concern is only deep-vein phlebitis, external pneumatic compression may be an acceptable alternative.

Anticoagulants can usually be safely discontinued for several days in patients being chronically maintained on them after myocardial infarction. Very high risk patients, such as those with artificial heart valves (particularly caged ball and disk valves in the mitral position), those with auricular fibrillation and enlarged left atria, those with a past history of phlebitis and thromboembolic episodes, and those with hypercoagulability states (antithrombin 3, protein S and C deficiencies, and the lupus anticoagulant syndrome), present major problems. The decision as to whether to completely correct them shortly before surgery and start heparin as early as possible postoperatively (12–48 hr) or to operate at a low level of anticoagulation (low-dose heparin or sufficiently reduced warfarin dosage to render the prothombin time 2–3 sec above control) must be arrived at on an individual basis and in collaboration with the surgical team. Even in the highest-risk patients (such as those with caged ball valves), a reasonably safe approach consists of discontinuation of warfarin several days prior to surgery, with simultaneous institution of full doses of intravenous heparin, which, in turn, is discontinued 6–12 hr preoperatively and resumed as early in the postoperative period (24 hr) as hemostasis permits. Warfarin is reinstituted some 3–4 days later.

If the decision is made to operate with the patient fully corrected, the following should be borne in mind:

- The half-life of heparin is about 90–100 min. Generally, the activated partial prothrombin time will return to acceptable levels within 4–6 hr after the last intravenous dose (less predictable with the irregularly absorbed subcutaneous route). If a more rapid effect is required, protamine, given diluted and slowly intravenously, in doses of 1 mg/100 units of heparin administered, is sufficient to accomplish normalization within minutes. Repeat dosing may be necessary, as the drug is cleared from the blood more rapidly than the heparin.
- Warfarin has a serum half-life of over 24 hr and the drug must be discontinued some 3–4 days earlier if the patient is to be operated on with a normal prothombin time. Vitamin $K_1$ given orally, subcutaneously, or slowly intravenously in doses of 2–20 mg, will have an impact on the prothombin time in 3–4 hr. Vitamin K takes somewhat longer. If there is overshoot, resistance to reanticoagulation results. In more urgent situations, fresh frozen plasma works rapidly, does not interfere with reanticoagulation, and replaces proteins C and S, but exposes the patient to the risks of blood products.

### Postoperative Factors

In the postoperative period, several additional factors are noteworthy.

First, hypercoagulability is common in the early postoperative period.

Second, if the patient has received large quantities of bank blood intraoperatively, minor hypocoagulability problems may develop. These are related to anticoagulant in the transfused blood, dilutional effects, and lack of replacement of blood-clotting factors and platelets.

Third, postoperative minidose heparin appears to be ineffective in preventing thromboembolic events in surgery of the hips, lower extremities, and in prostatectomy.

Fourth, in premenopausal women, estrogens are associated with a fivefold increase in the incidence of postoperative thromboembolism and should be discontinued 3–4 weeks prior to surgery.

### Suggested Reading

Rao TK, et al. Reinfarction following anesthesia in patients with myocardial infarction. Anesthesiology 1983; 59:499.

Dripps RD, et al. The role of anesthesia in surgical mortality. JAMA 1961; 178:261.

Marx GF, et al. Computer analysis of post-anesthetic deaths. Anesthesiology 1961; 39:54.

Wilson ME, et al. Assessment of fitness for surgical procedures and the variability of anesthetists' judgements. Br Med J 1980; 1:599.

Patterson KR. et al. The pre-operative ECG: an assessment. Scot Med J 1983; 28:116.

Robbins J, Mushlin A. Preoperative evaluation of the healthy patient. Med Clin North Am 1979; 63:1145.

Rucker L, et al. Usefullness of screening chest roentgenograms in pre-operative patients. JAMA 1983; 50:3209.

Seymour DG, et al. The role of routine pre-operative chest x-ray in the elderly general surgical patient. Postgrad Med J 1982; 58:741.

Kaplan E, et al. The usefulness of pre-operative laboratory screening. JAMA 1985; 253:3576.

Cygan R, Waitzkin H. Stopping and restarting medications in the perioperative period. J Gen Int Med 1987; 2:270.

Goldman I, Caldera DL. Risks of general anesthesia and elective operation in the hypertensive patient. Anesthesiology 1979; 50:285.

Katz RL. Hazardous effects of drugs in hypertensive patients scheduled for surgery. Cardiovasc Med 1978; 3:1185.

Goldman L, et al. Multifactorial index of cardiac risk in noncardiac surgical procedures. N Engl J Med 1977; 297:845.

Goldman L, et al. Cardiac risk factors and complications in non-cardiac surgery. Medicine 1978; 57:357.

Consensus conference: prevention of venous thrombosis and pulmonary embolism. JAMA 1986; 256:744.

Hull RD, et al. Different intensities of oral anticoagulant therapy in the treatment of proximal vein thrombosis. N Engl J Med 1982; 307:1676.

Katholi RE, Nolan SP, McGuire LB. The management of anticoagulation during noncardiac operations in patients with prosthetic heart valves: a prospective study. Am Heart J 1978; 96:163.

Francis CW, et al. Two step warfarin therapy. JAMA 1983; 249:374.

Gallus A, et al. Venous thrombosis after elective hip replacement. The influence of preventive intermittent calf compression and/or surgical technique. Br J Surg 1983; 70:17.

Turpie AGG, et al. Prevention of venous thrombosis in patients with intracranial disease by intermittent pneumatic compression boots. Surgery 1977; 83:354.

Schwartz AJ, Wollman H. Anesthetic considerations for patients on chronic drug therapy: L dopa, monoamine oxidase inhibitors, tricyclic antidepressants and propanalol. Anesthesiology 1976; 4:98.

# Anesthesia: What the Internist Needs to Know

Andrew B. Leibowitz, M.D., and Jeffrey H. Silverstein, M.D.

PERIOPERATIVE consultation intertwines anesthesiology, medicine, and surgery. Its goal is the efficient transfer of information in each consultant's specialty to the other members of the team caring for the patient, thus increasing the likelihood of a satisfactory outcome. Unfortunately, too often, this does not occur. The primary goal of this paper is to familiarize the internist with the principles of anesthesia, its risks and complications, so that effective, knowledgeable consultation may be possible.

## The Consultation

Studies report that

- 14% of medical subspecialty consultants disagreed with the requesting physician concerning the reason for the consultation (1)
- noncompliance by the requesting physician with recommendations ranges from 23% to 46% (2–4)
- compliance by the requesting physician decreases as the number of recommendations increases (4, 5)

In one recent study, 40% of requests for consultation by anesthesiologists were for "medical clearance," yet it is obvious that no one can guarantee a good outcome or ensure that the planned surgery and anesthetic will be safe (6).

"Clearance" is not a reasonable goal of consultation. The consultant should be asked a question; most often the question will involve an assessment of risk and a strategy to reduce the risk to a minimum in the perioperative period. Trivial suggestions like "avoid hypotension" are best avoided.

Goldman, whose name is closely linked to risk assessment, offers the following ten commandments for consultation (7), which serve as effective guidelines:

1. **Determine the question**
2. **Establish urgency**
3. **Look for yourself**
4. **Be as brief as appropriate**
5. **Be specific**
6. **Provide contingency plans**
7. **Honor thy turf (or, thou shalt not covet they neighbor's patient)**
8. **Teach . . . with tact**
9. **Talk is cheap . . . and effective**
10. **Follow up**

## Types of Anesthesia

General anesthesia is, in actuality, a pharmacologically induced, reversible coma. It is a state of suspended animation that seeks to attain five major goals: absolute analgesia; a controlled environment conducive to the surgery at hand; control of the physiologic responses to surgical stimulation; amnesia for the perioperative events; and avoidance of the toxic effects of the anesthetic agents. The general anesthetic state is a continuum beginning with changes in sensorium, progressing through planes of surgical anesthesia, and eventually leading to death if not properly controlled.

A wide variety of drugs are available for inducing and maintaining anesthesia. The potent volatile anesthetics (halothane, enflurane, isoflurane) are introduced to the body through the lungs and in themselves can create deep planes of

---

From the Department of Anesthesiology, The Mount Sinai Medical Center, New York, NY. Address reprint requests to Andrew B. Leibowitz, M.D., Box 1062, Department of Anesthesiology, Mount Sinai School of Medicine, One Gustave L. Levy Place, New York, NY 10029.

anesthesia consistent with the above goals. The opioids (morphine, fentanyl, meperidine), in large doses, may provide profound analgesia and stable physiologic signs but require additional medications to provide muscle relaxation and to avoid vivid recall of intraoperative events. Short-acting and long-acting barbiturates, benzodiazepines, antiemetics, and muscle relaxants are often used in addition. Rarely will "general anesthesia" include less than four or five medications. The choice of agents, as well as knowledge of their interactions with each other and with any medications a patient may habitually be taking, is the science and art of anesthesiology.

Regional anesthesia is, as the name implies, anesthesia for a region of the body. It is usually achieved by blocking nerve transmission from a given area with local anesthetics (for example, lidocaine). This blockade may be accomplished by injecting medication into the subarachnoid (spinal) or epidural space, or by infiltrating the area around specific nerve trunks (for instance, the brachial plexus). Regional anesthesia will provide total analgesia and good surgical conditions in specific body regions. Without supplemental medication the patient will be totally conscious of all intraoperative events. Complications associated with the use of regional anesthesia include hypotension, nerve injury, seizures, and cardiac arrest.

**The Controversy: Regional versus General Anesthesia.** The medical consultant should refrain from recommending a type of anesthesia. Recommending a specific anesthetic puts the anesthesiologist in a problematic position for three major reasons: the anesthesiologist may disagree with the choice; the anesthesiologist may be unable to use the recommended anesthetic for reasons of patient desire, the patient's anatomy, or the anesthesiologist's professional experience; no significant body of evidence suggests that one type of anesthesia is best for a particular situation (there are rare exceptions). Recently, for example, diametrically opposite viewpoints have been expressed on whether general or regional anesthesia should be used in patients with heart disease (8, 9). The choice of the anesthetic for a particular patient should therefore remain with the anesthesiologist who will be responsible for supervising its administration.

The current controversy over whether general or regional anesthesia directly results in improved outcome is based largely on poorly designed studies that, in sum, reach no firm conclusion. Only one major prospective study, involving 53 patients, reports a statistically significant improvement in outcome of high-risk patients when regional anesthesia was used (10). However, this study had many design flaws, including major variations in the technique of administration of the anesthetics used in the study and in addition an unusually high complication rate in those patients who received general anesthesia.

Regional anesthesia is accepted as superior to general anesthesia in certain circumstances: spinal anesthesia for transurethral resection of the prostate; epidural anesthesia for labor, delivery, and cesarean section; and local anesthesia for eye surgery. However, even these procedures may be safely performed using general anesthesia. Although regional anesthesia is considered by many to be superior to general anesthesia for patients with severe pulmonary disease, this superiority has not been established in clinical trials.

A recent retrospective study suggests that regional anesthesia may actually be associated with greater morbidity and mortality than is general anesthesia (11). There is little doubt that an expertly administered anesthetic of either type will yield a better outcome than a poorly administered anesthetic.

## Monitoring in the Operating Room

The anesthesiologist continuously observes and processes objective and subjective information directly from the anesthetized patient. Additional data are obtained from a variety of closely observed monitoring devices. Interpretation of data from all sources, early recognition of critical events (12), and timely intervention are essential elements of the art and science of anesthesia.

A set of standards for basic intraoperative monitoring has been published by the American Society of Anesthesiologists (13). A basic level of monitoring encompasses evaluation of oxygen in the inspired gas as well as blood oxygenation determined by pulse oximetry. An end-tidal carbon dioxide monitor also assesses the adequacy of ventilation. The heart and circulation are monitored by continuous electrocardiogram and frequent blood pressure determinations. Hourly urine output is commonly observed, especially if the procedure is expected to last more than 4 or 5 hr, and body temperature is measured and recorded.

Many operating rooms are equipped with either mass spectrophotometric or infrared analyzers, which can be used to rapidly determine the inhaled and exhaled concentrations of $O_2$,

$CO_2$, nitrogen, $N_2O$, and other anesthetic gases. The insertion of an arterial catheter, when indicated, makes possible beat-to-beat measurement of the arterial blood pressure and also allows frequent sampling of blood for analysis in the laboratory. Central venous catheters and pulmonary artery catheters are frequently employed to evaluate central venous pressure and cardiac function. Increasingly, intraoperative transesophageal echocardiography is being utilized to monitor patients at high cardiac risk.

Neurologic function can be evaluated by electroencephalogram (EEG) and processed EEG. The usefulness of these modalities is limited to specific situations (14). Peripheral neurologic function may be monitored by measurement of sensory, auditory, and visual evoked potentials.

The choice of what monitoring is needed is made by the anesthesiologist, who must consider the patient's underlying medical status, the type and length of the operation, and the patient's expected early postoperative course. The consultant can assist the anesthesiologist by clarifying the patient's past and current medical status, including details of therapeutic interventions. The anesthesiologist also will seek to elicit this information directly from the patient preoperatively whenever possible. The consultant who will actively participate in the postoperative care of the patient may suggest modalities that will be of assistance in patient care following the operation, because such suggestions may influence the anesthesiologist's selection of monitors.

## Complications of Anesthesia

Complications may arise in relation to any aspect of anesthesia and range from benign and short-lived problems to devastating complications, including death. Some of the major significant complications of interest to the internist include the following.

### Common "Benign" Complications

The most common complications of general anesthesia include nausea, vomiting, sore throat, and dental injuries. Corneal abrasion occasionally occurs if eye protection is not provided. Soft-tissue and nerve-compression injuries may result from improper positioning. Spinal headache is a common complication of spinal anesthesia performed with larger-bore needles (20 and 22 gauge) in younger patients and may also result from epidural anesthesia complicated by inadvertent dural puncture ("wet" tap). Regarding this complication, the internist should be aware that a blood patch, the injection of approximately 10 mL of autologous blood into the epidural space, is a safe, effective treatment for spinal headaches.

### Serious Intraoperative Complications

**Aspiration.** Prior to elective surgery, every adult patient takes no food or water after the midnight preceding scheduled surgery, to minimize the potential for aspiration of gastric contents and thus avoid compromise of airway management during anesthesia. In the operating room, aspiration occurs when an obtunded or unconscious patient passively regurgitates or actively vomits. Prevention of aspiration involves identification of patients at increased risk (those with hiatal hernia, diabetic gastroparesis, acute abdomen, obesity), reduction of the intragastric volume and gastric acidity, and successful protection of the airway. Aspiration of gastric contents with a pH below 2.5 and a sufficiently large volume (25 cc) can cause acute lung injury (15). If the status of gastric contents cannot be ascertained, consideration should be given to emptying the stomach, which is not without risk. Patients at risk are variably treated with a combination of H2 blockers, metoclopramide, nonparticulate antacids, and special airway manipulation.

**Esophageal Intubation.** The other major complication associated with control of the airway is unrecognized esophageal intubation (16). Control of the airway during anesthesia frequently requires placement of an endotracheal tube, which is usually but not universally accomplished without difficulty. Many potentially difficult intubations are clinically obvious from the outset and are approached as such; they can be performed with the patient awake and with fiberoptic guidance. More problematic are those cases in which the trachea proves unexpectedly difficult to intubate, particularly when the esophagus is accidentally intubated. Placement of the endotracheal tube in the esophagus is not in itself dangerous; failure to identify the problem and remedy the situation in an expedient fashion may lead to devastating hypoxic injury. A variety of monitors that detect expired $CO_2$ have improved immediate diagnosis of esophageal intubation.

**Death.** The Harvard Affiliated Hospitals recently reported the results of 1,329,000 anesthetics administered from 1976 through mid-1988 (17). Among 1,001,000 American Society of Anesthesiologists physical status I and II patients,

TABLE 1
*Goldman Risk Index*

| Condition | Points |
|---|---|
| Congestive heart failure (includes S3, or JVD) | 11 |
| MI < 6 months ago | 10 |
| PVC > 5/min | 7 |
| Rhythm other than sinus | 7 |
| Age >70 | 5 |
| Emergency operation | 4 |
| Intrathoracic, intraperitoneal, or aortic surgery | 3 |
| Aortic stenosis | 3 |
| Poor overall condition | 3 |

| Class | Points | Complication rate | Cardiac death rate |
|---|---|---|---|
| I | 0–5 | <1% | 0 |
| II | 6–12 | 5% | 2% |
| III | 13–25 | 11% | 2% |
| IV | >25 | 78% | 56% |

11 major intraoperative accidents were solely attributable to anesthesia (5 deaths, 4 cases of permanent central nervous system damage, and 2 cardiac arrests with eventual recovery). Seven of these accidents occurred as a direct result of unrecognized hypoventilation. To put these data into perspective, they suggest that the risk of a relatively healthy person dying from a primary anesthetic error is less than the risk of being a homicide victim in a one-year period in the city of New York. The relatively recent widespread acceptance of capnometry and pulse oximetry may even lessen this risk (18). In fact, only one of the reported incidents in the Harvard study occurred after the 1985 adoption of routine monitoring standards. Previous studies performed in the era before capnometry and pulse oximetry reported anesthesia death rates as high as 3.7/10,000 anesthetics in 1948–1952 (19), and ranging to 0.7–2.2/10,000 anesthetics administered in the subsequent 30-year period (20).

TABLE 2
*Postoperative Myocardial Infarction in Patients with Previous Myocardial Infarction*

| Months post-MI | 1973–1976 % | 1973–1976 No. | 1976–1982 % | 1976–1982 No. |
|---|---|---|---|---|
| 0–3 | 30% | 4/11 | 5.8% | 3/52 |
| 4–6 | 26% | 8/31 | 2.3% | 2/36 |
| 7–12 | 5% | 6/127 | 1.0% | 1/104 |
| >12 | 5% | 10/19 | 1.7% | 8/489 |

Adapted from Rao TL, Jacobs KH, El Etr AA. Reinfarction following anesthesia in patients with myocardial infarction. Anesthesiology 1983; 59:499–505 (ref 24).

**Cardiac Complications.** Patients at low risk rarely have postoperative cardiac complications. In fact, myocardial infarction occurs in less than 1% of all patients undergoing surgery (21). Risk is most often assessed by the Goldman Risk Index (Table 1) (22, 23). Recent myocardial infarction and even minor symptomatic or objective evidence of congestive heart failure are the most important predictors of postoperative complications. The Goldman criteria were derived for all patients undergoing noncardiac surgery; patients undergoing vascular surgery have a higher complication rate. However, since Goldman's data were first published, in 1977, complication rates in at least one practitioner's experience have markedly improved.

Rao et al. (Table 2) compared the rates of postoperative myocardial infarction during two periods of time, showing a 5.8% and 2.3% rate of reinfarction in patients less than 3 and 6 months postinfarction, respectively, in the years 1976–1982 (24). Previous studies reported reinfarction rates consistent with Rao's data for 1973–1976 (21, 25). This improvement was attributed to invasive monitoring, intensive hemodynamic manipulation, and postoperative care in intensive care units. More recently, Rao has also reported a reduction in cardiac mortality to 8.2% and in morbidity to 18.4% in Goldman Class IV patients (26).

Current work has emphasized the proper preoperative evaluation of the high-risk patient. Persantine-thallium imaging appears to easily identify a group of patients at particularly high risk (27). The challenge remains to improve their outcome by a different anesthetic or operative approach or more invasive preoperative study and possible coronary artery bypass grafting. Most recently, Holter monitoring for ischemia has been shown to aid in risk stratification and may further guide the consultant to further studies and treatment prior to surgery (28).

Consultants need to be aware that postoperative myocardial infarction differs from myocardial infarction occurring in the outpatient who comes to the hospital with chest pain. Postoperative myocardial infarction is usually painless, occurs more commonly three and four days postoperatively, rather than on the first postoperative day, and carries a significantly worse prognosis, entailing a 50% mortality rate (29, 30).

**Hepatitis.** Postoperative hepatitis is rarely a result of anesthesia, though volatile halogenated agents have been associated with hepatitis since the first description of chloroform in 1847 (31). Halothane, the volatile agent most closely linked to hepatic injury, may be responsible for postoperative hepatitis in 1/10,000–20,000 administrations (32, 33). Of note, this is far less than the rate of undiagnosed viral hepatitis in patients who come for elective surgery (34). Factors predisposing to the development of "halothane hepatitis" include multiple exposures and obesity; women are more predisposed than men (32, 35). Since the development of enflurane and isoflurane, the majority of adults undergoing general anesthesia do not receive halothane; postoperative hepatitis resulting from an anesthetic is thus extremely rare. Enflurane hepatotoxicity may exist, possibly occurring in 1/1,000,000 administrations (36). Isoflurane seems even less likely to cause postoperative hepatitis (37).

**Malignant Hyperthermia.** Malignant hyperthermia is an extremely rare inherited disorder involving abnormal calcium release and uptake in the muscle, triggered by the administration of the short-acting depolarizing muscle relaxant succinylcholine or one of the potent inhalational anesthetics. The incidence of the disorder in adults is approximately 1/50,000; in children it is probably more frequent. Typically, the patient develops tachycardia, arrhythmias, tachypnea or pharmacologic paralysis, elevated end-tidal $CO_2$ and arterial $pCO_2$, and severe hyperthermia. Treatment consists of early discontinuation of the offending agent, body cooling, intravenous fluid administration, and, most important, intravenously administered dantrolene. Subsequent anesthesia performed without succinylcholine and potent volatile agents is safe. Patients with affected family members, or certain inherited muscle diseases, are at risk and should undergo muscle biopsy for halothane caffeine contracture testing to determine their susceptibility. Physicians caring for these patients must be aware of the potentially fatal nature of the condition and keep abreast of the rapidly changing recommendations for diagnosis and treatment (38). The Malignant Hyperthermia Association of the United States (MHAUS) publishes a quarterly newsletter and staffs a 24-hour hotline to answer physicians' questions (209-634-4917).

**Renal Failure.** Most anesthetic agents cause a transient decrease in glomerular filtration rate and in urinary output. However, renal failure is not a consequence, unless prolonged hypotension and hypovolemia occur (39). Fluoride-induced renal failure occurred with the use of methoxyflurane, an agent no longer used (40). Significant nephrotoxicity may occur from high fluoride levels following prolonged exposure to high concentrations of enflurane, especially if preexisting renal dysfunction exists (41). However, in a report on a large series of patients undergoing elective surgery with enflurane anesthesia, some improvement in renal function was noted (42). In any case, isoflurane may be used safely in almost every clinical situation and the possibility of fluoride-induced nephrotoxicity entirely avoided.

## Underlying Disease

The consultant evaluating a patient prior to surgery needs to address not only the patient's general medical condition but also certain issues important to the anesthesiologist. The following, though not inclusive, is a synopsis of those clinical findings that often require evaluation.

**Lung Disease.** Surgical intervention in the patient with chronic lung disease carries a much increased risk of postoperative respiratory failure (43). Preoperative consultation should focus on the nature and degree of disease, its reversibility, and optimum management.

Although the anesthesiologist is aware of the implications of emphysema, chronic bronchitis, and acute asthma for safe anesthesia, evaluation and optimization of overall pre- and post-operative care are best accomplished in conjunction with the internist-pulmonologist. The history, pulmonary function studies, and evaluation of arterial blood gases are particularly important if the disease is severe, particularly if the patient will be undergoing thoracic or upper abdominal surgery. It is known, for example, that a ratio of $FEV_1$:VC of less than 50% is associated with an increased risk of postoperative respiratory failure (44). Moreover, preoperative improvement in the patient's status is known to lessen the incidence of postoperative pulmonary complications (45).

Determination of lung resectability is a more specialized area and may require sophisticated split lung function tests and possibly unilateral pulmonary artery occlusion.

The presence of cor pulmonale invokes another set of risks and appropriate pre- and postoperative therapies for minimizing operative risk.

**Hepatitis and Cirrhosis.** The patient with active liver disease presents a unique set of problems for both consultant and anesthesiologist. Although the many studies assessing the risk of surgery in these patients are difficult to reconcile because of high variability in patient selection, it is clear that the operative morbidity and mortality of patients with liver disease is high and is probably related to the degree of hepatic dysfunction (46). One report describes 30% mortality among a group of 100 cirrhotic patients undergoing nonshunt abdominal surgery, 87% of which was due to sepsis with multiple organ system failure (47).

Particular attention must be paid to assessment and correction of coagulation disorders prior to surgery. In addition, any anesthetic agent may decrease liver blood flow and potentiate any underlying liver injury. Excellent reviews of this subject provide further information (48, 49).

**Stroke.** The patient who has recently had a stroke is frequently referred for preoperative consultation. When elective surgery is safe to perform is unclear. One of the primary concerns of the anesthesiologist is the return of normal cerebral autoregulation, so that the patient may tolerate the swings in blood pressure that often accompany surgery and anesthetic administration. Two weeks after an uncomplicated stroke has been suggested as a reasonable period of time to expect this return; however, this recommendation on timing surgery is based on scant data (50).

## Steroids

It is now clear that the cortisol response to operative stress varies by type of surgery (51). For minor surgery the patient may take 1½ to 2 times the usual prednisone dose on the morning of surgery, and the following day may take the regular prednisone dose. When major surgery is required, the patient should receive twice the regular daily glucocorticoid dosage on the morning of surgery, or 50 mg intravenous hydrocortisone preoperatively, then 100 mg intravenous hydrocortisone intraoperatively. Postoperatively, 100 mg hydrocortisone may be administered every 8 hours for 24 hours. Thereafter, if the postoperative course is uncomplicated, this high dose should be tapered rapidly toward the usual maintenance dose (52).

## Summary

We have attempted to acquaint the internist with some aspects of anesthesiology that need to be kept in mind when performing perioperative consultation. Communication among and between the entire operative team will reduce risk and untoward reactions and will enhance the likelihood of successful outcome and rapid recovery.

## References

1. Lee T, Pappius EM, Goldman L. Impact of inter-physician communication on the effectiveness of medical consultations. Am J Med 1983; 74:106–112.
2. Klein LE, Levine DM, Moore RD, Kirby SM. The preoperative consultation. Response to internists' recommendations. Arch Intern Med 1983; 143:743–744.
3. Ballard WP, Gold JP, Charlson ME. Compliance with the recommendations of medical consultants. J Gen Intern Med 1986; 1:220–224.
4. Sears CL, Charlson ME. The effectiveness of a consultation. Compliance with initial recommendations. Am J Med 1983; 74:870–876.
5. Mackenzie TB, Popkin MK, Callies AL, Jorgensen CR, Cohn JN. The effectiveness of cardiology consultation. Concordance with diagnostic and drug recommendations. Chest 1981; 79:16–22.
6. Kleinman B, Czinn E, Shah K, Sobotka PA, Rao TK. The value to the anesthesia-surgical care team of the preoperative cardiac consultation. J Cardiothorac Anesth 1989; 3:682–687.
7. Goldman L, Lee T, Rudd P. Ten commandments for effective consultations. Arch Intern Med 1983; 143:1753–1755.
8. Yeager MP. Pro: regional anesthesia is preferable to general anesthesia for the patient with heart disease. J Cardiothorac Anesth 1989; 3:793–796.
9. Beattie C. Con: regional anesthesia is not preferable to general anesthesia for the patient with heart disease. J Cardiothorac Anesth 1989; 3:797–800.
10. Yeager MP, Glass DD, Neff RK, Brinck Johnsen T. Epidural anesthesia and analgesia in high-risk surgical patients. Anesthesiology 1987; 66:729–736.
11. Manolia TA, Beattie C, Christopherson R, Pearson T. Regional vs general anesthesia in high risk surgical patients: need for clinical trial. J Clin Anesth 1989; 1:414.
12. Cooper JB, Newbower RS, Kitz RJ. An analysis of major errors and equipment failures in anesthesia management: considerations for prevention and detection. Anesthesiology 1984; 60:34–42.
13. Standards for basic intra-operative monitoring. FYI . . . A commitment to patient safety. Washington, DC: ASA Office of Governmental Affairs, 1990:8–9.
14. Donegan JH, Rampil IJ. The electroencephalogram. In: Blitt CD, ed. Monitoring in anesthesia and critical care medicine. New York: Churchill Livingstone, 1990: 431–459.
15. Wynne JW, Modell JH. Respiratory aspiration of stomach contents. Ann Intern Med 1977; 87:466–474.
16. Marx GF, Mateo CV, Orkin LR. Computer analysis of postanesthetic deaths. Anesthesiology 1973; 39:54–58.
17. Eichhorn JH. Prevention of intraoperative anesthesia accidents and related severe injury through safety monitoring. Anesthesiology 1989; 70:572–577.
18. Orkin FK. Practice standards: the Midas touch or the emperor's new clothes. Anesthesiology 1989; 70:567–571.
19. Beecher HK, Todd DP. A study of deaths associated with anesthesia and surgery based on 599,548 anesthetics in 10 institutions, 1948–1952 inclusive. Ann Surg 1954; 140:2–35.
20. Duperman SM, Bendixen HH. Mortality, morbidity and risk studies in anesthesia, epidemiology in anesthesia. In: Lunn JN, ed. The techniques of epidemiology applied to anesthetic practice. Baltimore: Edward-Arnold, 1986:37–73.
21. Topkins MJ, Artusio JF. Myocardial infarction and surgery: a five year study. Anesth Analg 1964; 43:716–720.

22. Goldman L, Caldera DL, Nussbaum SR, et al. Multifactorial index of cardiac risk in noncardiac surgical procedures. N Engl J Med 1977; 297:845–850.
23. Goldman L. Cardiac risks and complications of noncardiac surgery. Ann Intern Med 1983; 98:504–513.
24. Rao TL, Jacobs KH, El Etr AA. Reinfarction following anesthesia in patients with myocardial infarction. Anesthesiology 1983; 59:499–505.
25. Steen PA, Tinker JH, Tarhan S. Myocardial reinfarction after anesthesia and surgery. JAMA 1978; 239:2566–2570.
26. Shah K, Kleinman B, Rao T: Reduction in mortality from cardiac causes in Goldman class IV patients. J Cardiothorac Anesth 1988; 2:789–791.
27. Boucher CA, Brewster DC, Darling RC, Okada RD, Strauss HW, Pohost GM. Determination of cardiac risk by dipyridamole-thallium imaging before peripheral vascular surgery. N Engl J Med 1985; 312:389–394.
28. Raby KE, Goldman L, Creager MA, et al. Correlation between preoperative ischemia and major cardiac events after peripheral vascular surgery [see comments]. N Engl J Med 1989; 321:1296–1300.
29. Goldman L, Caldera DL, Southwick FS, et al. Cardiac risk factors and complications in non-cardiac surgery. Medicine (Baltimore) 1978; 57:357–370.
30. Salem DN, Homan DC, Isner JM. Management of cardiac disease in the general surgical patient. Curr Probl Cardiol 1980; 5:1–41.
31. Zimmerman HJ. Hepatotoxicity: the adverse effect of drugs and other chemicals on the liver. New York: Appleton Century Crofts, 1978:370–394.
32. Inman WH, Mushin WW. Jaundice after repeated exposure to halothane: an analysis of Reports to the Committee on Safety of Medicines. Br Med J 1974; 1:5–10.
33. Bottiger LE, Dalen E, Hallen B. Halothane-induced liver damage: an analysis of the material reported to the Swedish Adverse Drug Reaction Committee, 1966–1973. Acta Anaesthesiol Scand 1976; 20:40–46.
34. Schemel WH. Unexpected hepatic dysfunction found by multiple laboratory screening. Anesth Analg 1976; 55:810–812.
35. Brown BR Jr, Gandolfi AJ. Adverse effects of volatile anaesthetics. Br J Anaesth 1987; 59:14–23.
36. Lewis JH, Zimmerman HJ, Ishak KG, Mullick FG. Enflurane hepatotoxicity. A clinicopathologic study of 24 cases. Ann Intern Med 1983; 98:984–992.
37. Stoelting RK, Blitt CD, Cohen PJ, Merin RG. Hepatic dysfunction after isoflurane anesthesia. Anesth Analg 1987; 66:147–153.
38. Rosenberg H, Seitman D. Pharmacogenetics. In: Barash PG, Cullen BF, Stoelting RK, eds. Clinical anesthesia. Philadelphia: Lippincott, 1989:459–484.
39. Charlson ME, MacKenzie CR, Gold JP, Ales KL, Shires GT. Postoperative renal dysfunction can be predicted. Surg Gynecol Obstet 1989; 169:303–309.
40. Cousins MJ, Mazze RI. Methoxyflurane nephrotoxicity. A study of dose response in man. JAMA 1973; 225:1611–1616.
41. Loehning RW, Mazze RI. Possible nephrotoxicity from enflurane in a patient with severe renal disease. Anesthesiology 1974; 40:203–205.
42. Mazze RI, Sievenpiper TS, Stevenson J. Renal effects of enflurane and halothane in patients with abnormal renal function. Anesthesiology 1984; 60:161–163.
43. Tarhan S, Moffitt EA, Sessler AD, Douglas WW, Taylor WF. Risk of anesthesia and surgery in patients with chronic bronchitis and chronic obstructive pulmonary disease. Surgery 1973; 74:720–726.
44. Stein M, Cassara EL. Preoperative pulmonary evaluation and therapy for surgery patients. JAMA 1970; 211:787–790.
45. Gass GD, Olsen GN. Preoperative pulmonary function testing to predict postoperative morbidity and mortality. Chest 1986; 89:127–135.
46. Lindenmuth WW, Eisenberg MM. The surgical risk in cirrhosis of the liver. Arch Surg 1963; 86:77–84.
47. Garrison RN, Cryer HM, Howard DA, Polk HC Jr. Clarification of risk factors for abdominal operations in patients with hepatic cirrhosis. Ann Surg 1984; 199:648–655.
48. Friedman LS, Maddrey WC. Surgery in the patient with liver disease. Med Clin North Am 1987; 71:453–476.
49. Tewes P. Liver disease, including postoperative hepatic dysfunction. In: Breslow MJ, Miller CF, Rogers M, eds. Perioperative management. St. Louis: Mosby, 1990: 404–416.
50. Merli GJ, Bell RD. Preoperative management of the surgical patient with neurological disease. Med Clin North Am 1987; 71:511–527.
51. Chernow B, Alexander HR, Smallridge RC, et al. Hormonal responses to graded surgical stress. Arch Intern Med 1987; 147:1273–1278.
52. Chernow B. Perioperative management of endocrine problems: thyroid, adrenal cortex, pituitary, ASA 1988 Annual Refresher Course Lectures. American Society of Anesthesiologists, 1988:1–5.

# The Art and Science of Consultation in Medicine

## Preoperative Prevention

# Preoperative Preventive Antimicrobial Therapy

Eric Neibart, M.D.

A General consensus has emerged among surgeons, internists, and infectious disease subspecialists that preoperative systemic antimicrobial prophylaxis for the prevention of surgical wound infection is highly efficacious. In a clean wound (no entry into the respiratory, gastrointestinal [GI], or genitourinary [GU] tract, no break in surgical technique) there is an incidence of infection of about 5%; in a clean-contaminated wound (entry into the respiratory, GI, or GU tract without contamination of the surgical field) there is a wound infection incidence of 8–10%; in a contaminated wound (open fresh traumatic wounds without pus, or major break in sterile technique or major GI spill), in the presence of inflammation (without pus) the incidence of infection is in the range of 20%, and in dirty wounds (established infectious process), the incidence is 40% or more (1, 2). The use of prophylactic antibiotics decreased the infection rates to 0.8%, 1.3%, and 10.2% in clean, clean-contaminated, and contaminated wounds, respectively (3). As a general rule, prophylactic antibiotics are not necessary in clean wounds unless there are special circumstances, such as a recognized risk of endocarditis or the surgery involves the implanting of a foreign body. Dirty wounds are generally treated as infections.

Factors that have an impact on the incidence of wound infection other than the type and site of operation and surgical technique include the length of surgery (>2 hr) and whether or not it is an emergency use of a heart-lung machine, age, and underlying medical conditions (malignancy, diabetes, obesity, malnutrition, steroid administration and immunoincompetency, and local vascular disease). The placement of a foreign body (central nervous system shunt, bone or valvular prosthesis, arterial or venous graft) likewise increases the likelihood of and the catastrophic result of infection. The existence of an elevated pH ($H^2$ blockers, antacids) predisposes to infection in gastroduodenal surgery, as does obstruction in biliary and bowel procedures. Central lines and indwelling urethral and arterial catheters also increase the potential for infection, and their early removal is itself prophylactic. Urinary tract and pulmonary infections should be treated and controlled prior to surgery.

Two concepts underlie the current approach to prophylaxis:

First, despite the use of routine aseptic techniques, it was demonstrated that *Staphylococcus aureus* could regularly be isolated from the operative field, indicating that aseptic technique reduced, but did not eliminate, bacterial colonization or contamination of the surgical wound (4–6).

The second development revealed a flaw in the design of all previous studies which failed to reveal a benefit of antimicrobial prophylaxis. In these earlier clinical studies, antibiotics were administered only after the surgery. Using a guinea-pig model of infection, Burke demonstrated that the administration of penicillin shortly before or at the time of intradermal inoculation of *S. aureus* into the animal resulted in a significant reduction of the subsequent wound infection (7, 8). But if the administration of antibiotics was delayed only several hours, the size of the wound infection was no different than in the control animals. This study led to the conclusion

---

From the Division of Infectious Diseases, Department of Medicine, Mount Sinai School of Medicine, New York, N.Y. 10029. Address reprint requests to the author at 1112 Park Avenue, New York, NY 10029.

that, for antimicrobial prophylaxis to be effective, the antimicrobial agent must be in the tissue at the time of bacterial contamination.

The benefits of prophylaxis prior to contamination have now been established by numerous investigators in a variety of experimental models and hundreds of clinical trials. In a review of the literature, Guglielmo et al. (9) identified more than 150 studies that met accepted criteria for assessment of an adequate trial and found that 80% demonstrated a significant benefit ($p < 0.05$) of prophylaxis (Tables 1, 2) (8, 10). It is now established that antibiotics are most likely to prevent infection when the number of organisms is low, when they are multiplying rapidly, when the site is not yet deprived of blood, and when antibiotics are in tissues at the time of exposure to bacteria.

There remain many unresolved issues, including optimal choice of an antimicrobial agent, timing and frequency of administration, duration of perioperative and postoperative prophylaxis, and potential complications.

## Choice of Antibiotic

The choice of antibiotic is dictated by the site of surgery and type of flora anticipated as well as by any variations in the type or resistance of the nosocomial pathogens peculiar to the hospital at which the procedure is being performed. Protocols and drugs of choice are multiple and change constantly as new studies and new antibiotics become available. Excellent summaries of current recommendations for the various procedures are available (Tables 3, 4) (10, 11). Clinical studies to detect differences between a variety of antimicrobial agents are difficult in view of the large sample size (>900) required to detect a significant ($p = 0.05$) difference (12, 13).

In general, the cephalosporin antibiotics have been the most widely utilized agents for prophylaxis based on their antibacterial spectrum and low incidence of allergy and toxicity.

Cefazolin, by virtue of its modest cost, broad antibacterial spectrum, and relatively long half-life, has become the dominant agent for clean and

**TABLE 1**

*Classification of Operative Wounds by Level of Bacterial Contamination*

| | % infection rate | |
|---|---|---|
| | Preoperative antibiotics routinely administered | |
| Classification criteria | No* | Yes† |
| **Clean wound**<br>nontraumatic wound in which no inflammation was encountered, no break in technique occurred, and respiratory, alimentary, and genitourinary tracts were not entered | 5.1 | 0.8 |
| **Clean-contaminated wound**<br>nontraumatic wound in which a minor break in technique occurred or in which the gastrointestinal, genitourinary, or respiratory tracts were entered without significant spillage; transection of appendix or cystic duct in the absence of acute inflammation and entrance into biliary or genitourinary tracts in the absence of infected bile or urine | 10.1 | 1.3 |
| **Contaminated wound**<br>fresh, traumatic wound from a relatively clean source or an operative wound in which there is a major break in technique, gross spillage from the gastrointestinal tract, or entrance into the genitourinary or biliary tract in the presence of infected urine or bile; incisions encountering acute nonpurulent inflammation; dirty wounds, such as traumatic wounds from a dirty source or with delayed treatment, fecal contamination, foreign bodies, a devitalized viscus, or pus from any source that is encountered | 21.9 | 10.2 |

* Data obtained from a multicenter study performed at a time (1960–1961) when antibiotic "prophylaxis," if administered, was usually initiated postoperatively. Classification criteria in 1960–1961 were less restrictive in that cholecystectomies, appendectomies, hysterectomies, and urinary tract operations were included as "clean" if no inflammation was present at the time of operation. Contaminated wounds were also subdivided into "contaminated" and "dirty"; the latter included old traumatic wound and those wounds involving abscesses or perforated viscus.

† Data obtained from a prospective single-center study of 20,000 wounds in which a fall in infection rates over a 5-year period was attributable to an increasing use of preoperative antibiotics.

*Adapted* from Kaiser (ref 8).

clean-contaminated procedures in which aerobic bacteria are encountered (10, 14). It is useful in upper gastrointestinal, biliary, obstetric-gynecologic (in the absence of pelvic inflammatory disease), head and neck, arterial (brachial and cephalic vessel surgery usually require no prophylaxis), orthopedic (especially joint replacement and internal fixation procedures), thoracic, and plastic procedures.

In the presence of allergy of IGE type to penicillin or cephalosporin, individualization of therapy is required. Depending on the expected flora, vancomycin with or without an aminoglycoside or the combination of clindamycin and an aminoglycoside might be appropriate.

Although cefazolin is still recommended by many, recent data revealing that it fails to prevent *S. aureus* infection in cardiac surgery have led to a consideration of alternatives (vancomycin and possibly cefmandole) in situations where *S. aureus* is the principal pathogen (15).

In surgical procedures in areas in which anaerobes are expected (appendix, gangrene), cefoxitin or cefotetan provides excellent prophylaxis. If there is penicillin allergy, clindamycin or metronidazole may be used with an aminoglycoside.

Oral neomycin and erythromycin given in the 24 hr prior to elective colorectal surgery provide excellent antibiotic prophylaxis for such procedures. When surgery is more urgent, cefoxitin is effective intravenously. The benefit of using combined oral and intravenous therapy is unclear and requires further investigation.

Dirty surgery, such as for a perforated viscus, is usually managed with cefotoxin, cefotetan, or a combination of an aminoglycoside and clindamycin or metronidazole plus ampicillin.

The efficacy of antibiotic prophylaxis has not been established in neurosurgery. Trials with trimethoprim-sulfamethoxazole, nafcillin, cefazolin, clindamycin, vancomycin, and local irrigations with the latter two drugs have not been definitive, although most neurosurgeons utilize one or more of them. Cefazolin is adequate prophylaxis for most ear, nose, and throat surgery, but in invasive or contaminated surgery, clindamycin with or without gentamicin may be efficacious. The urine should be rendered bacteria-free prior to prostatic surgery.

Any dental, surgical, or endoscopic procedure with biopsy to be performed in patients with valvular heart disease (including idiopathic hypertropic subaortic stenosis and mitral valve prolapse with a regurgitant murmur), most congenital heart lesions, a history of bacterial endocarditis, an artificial heart valve, or a surgically constructed pulmonary shunt should be preceded by antibiotic prophylaxis for endocarditis. Efficacy trials have not been performed in humans, and case reports are available describing failures (16). However, none of the failures reported in the paper by Durack et al. (16) received the current recommended antibiotic doses or timing of administration.

TABLE 2

*Criteria for Designing or Assessing Studies of Antimicrobial Prophylaxis of Surgical Wound Infection*

1. The study is prospective, randomized, and double blind.
2. Written definitions of surgical wound infections are established.
3. Comparability of host risk factors is present in all arms of the study.
4. The various operative procedures are well defined and equally distributed in all arms of the study. The descriptions of surgical procedures should be as focused as possible—e.g., "abdominal-peritoneal resection" should be used instead of "colon surgery," and "carotid endarterectomy" instead of "vascular surgery."
5. The choice of antimicrobial is appropriate for the expected pathogens and their antimicrobial susceptibilities.
6. The doses of the antimicrobial agents, as well as the initiation, route, frequency, and duration of their administration, are chosen to provide adequate tissue levels throughout the operative procedure.
7. Concomitant use of nonstudy perioperative antibiotics and antiseptics is controlled or eliminated.
8. The antimicrobials used carry a minimum risk of toxic or allergic reactions.
9. The costs of the various approaches, including the cost of any infections, are compared.
10. Sufficient numbers of patients are included to demonstrate whether significant differences do or do not exist between antibiotic regimens.

*From* Kaiser (ref. 10).

The Committee on Rheumatic Fever and Infectious Carditis (17) recommends, for dental and respiratory procedures, penicillin V, 2 g orally 1 hr prior to the procedure and 1 g 6 hr later. If penicillin allergy exists, erythromycin given orally, 1 g 2 hr prior to the procedure and 500 mg 6 hr later, suffices. In high-risk patients (artificial valves and pulmonary shunts), maximum protection is afforded by parenteral administration of 2 g of ampicillin and 1.5 mg/kg of gentamicin 30 min prior to the operation and 1 g of penicillin V orally 6 hr later. When penicillin allergy exists, vancomycin may be used. If gastrointestinal or genitourinary surgery is contemplated, standard prophylaxis mandates parenteral ampicillin and gentamycin as above and perhaps a repeat dosage 8 hr later. More recent advice (18) suggests that oral amoxicillin, 3 g given 1 hr prior to surgery and 1.5 g 6 hr later, is adequate in this circumstance. Penicillin-aller-

TABLE 3

*Prevention of Wound Infection and Sepsis in Surgical Patients*

| Nature of operation | Likely pathogens | Recommended drugs | Adult dosage before surgery* |
|---|---|---|---|
| Clean | | | |
| Cardiac | | | |
| Prosthetic valve and other open-heart surgery | *Staphylococcus epidermidis*, *S. aureus*, *Corynebacterium*, enteric gram-negative bacilli | Cefazolin[†]<br>or vancomycin[‡] | 1 g IV<br>1 g IV |
| Vascular | | | |
| Arterial surgery involving the abdominal aorta, a prosthesis, or a groin incision | *S. aureus*, *S. epidermidis*, enteric gram-negative bacilli | Cefazolin<br>or vancomycin[‡] | 1 g IV<br>1 g IV |
| Lower extremity amputation for ischemia | *S. aureus*, *S. epidermidis*, enteric gram-negative bacilli, clostridia | Cefazolin | 1 g IV |
| Orthopedic | | | |
| Total joint replacement, internal fixation of fractures | *S. aureus*, *S. epidermidis* | Cefazolin<br>or vancomycin[‡] | 1 g IV<br>1 g IV |
| Ocular[§] | *S. aureus*, *S. epidermidis*, streptococci, enteric gram-negative bacilli, *Pseudomonas* | Gentamicin or tobramycin or combination of neomycin, gramicidin, and polymyxin B | Multiple drops topically over 2–24 hr |
| Clean-contaminated | | | |
| Head and neck | | | |
| Entering oral cavity or pharynx | *S. aureus*, streptococci, oral anaerobes | Cefazolin | 2 g IV[‖] |
| Gastroduodenal | Enteric gram-negative bacilli, gram-positive cocci | High risk, gastric bypass, or percutaneous endoscopic gastrostomy only: cefazolin | 1 g IV |
| Biliary tract | Enteric gram-negative bacilli, enterococci, clostridia | High risk only: cefazolin | 1 g IV |
| Colorectal | Enteric gram-negative bacilli, anaerobes | Oral: neomycin plus erythromycin base | 1 g of each at 1 P.M., 2 P.M., and 11 P.M. the day before the operation[#] |
| | | Parenteral: cefoxitin | 1 g IV |

gic patients should receive 1 g of vancomycin and gentamicin (1.5 mg/kg), both given slowly IV (over an hour) prior to surgery; this may be repeated 8 hr later. Minor or repeated procedures in low-risk patients are adequately covered by 3 g of amoxicillin orally 1 hr prior to and 1.5 g 6 hr after the procedure.

### Timing Administration of Antibiotic Agents

There is general agreement that the initial dose of the antimicrobial agent should be given shortly prior to the procedure. The administration of "on-call" antibiotics is discouraged because unexpected delays in surgery may lead to subtherapeutic levels late in the operation (19). Repeated doses (every 4–8 hr) are required in prolonged procedures to maintain adequate therapeutic tissue levels throughout. This has been demonstrated in a group of patients undergoing cardiac surgery, in whom a higher infection rate was noted in those whose serum antibiotic level was undetectable at the conclusion of surgery (20).

### Duration of Postoperative Prophylaxis

How long prophylaxis should continue postoperatively remains controversial. Most studies comparing short- and long-course prophylaxis have failed to reveal any increase in infection among the short-course recipients (Table 5) (8). Lack of documented benefit of prolonged duration of therapy, its increased cost, and the potential for bacterial resistance lead to the current preference for short-duration therapy.

TABLE 3

*Prevention of Wound Infection and Sepsis in Surgical Patients (continued)*

| Nature of operation | Likely pathogens | Recommended drugs | Adult dosage before surgery* |
|---|---|---|---|
| Appendectomy | Enteric gram-negative bacilli, anaerobes | Cefoxitin | 1 g IV |
| Vaginal or abdominal hysterectomy | Enteric gram-negative bacilli, anaerobes, group B streptococci, enterococci | Cefazolin | 1 g IV |
| Cesarean section | Same as for hysterectomy | High risk only: cefazolin | 1 g IV after cord clamping |
| Abortion | Same as for hysterectomy | First trimester in patients with previous pelvic inflammatory disease: | |
| | | aqueous penicillin G | 1 million units IV |
| | | or doxycycline | 300 mg PO |
| | | Second trimester: cefazolin | 1 g IV |
| **Dirty** | | | |
| Ruptured viscus | Enteric gram-negative bacilli, anaerobes, enterococci | Cefoxitin with or without<br>gentamicin<br>or clindamycin plus<br>gentamicin | 1 g q6h IV<br>1.5 mg/kg q8h IV<br>600 mg IV q6h<br>1.5 mg/kg q8h IV |
| Traumatic wound** | *S. aureus*, group A streptococci, clostridia | Cefazolin | 1 g q8h IV |

* Parenteral prophylactic antimicrobials for clean and clean-contaminated surgery can be given as a single intravenous dose just before the operation. Cefazolin can also be given intramuscularly. For prolonged operations, additional intraoperative doses should be given q4–8h for the duration of the procedure. For "dirty" surgery, therapy should usually be continued for five to 10 days.

† Some infectious disease specialists prefer cefuroxime (Zinacef; Kefurox) or cefamandole (Mandol) for this indication (LD Sabath, Antimicrob Agents Chemother, 33:407, 1989).

‡ For hospitals in which methicillin-resistant *S. aureus* and *S. epidermidis* frequently cause wound infection, or for patients allergic to penicillins or cephalosporins.

§ In addition, at the end of the operation many ophthalmologists give a subconjunctival injection of an aminoglycoside such as gentamicin (10 to 20 mg), with or without a cephalosporin such as cefazolin (100 mg).

‖ In controlled studies, 2 g was effective, while 0.5 g was not (JT Johnson and VL Yu, Ann Surg, 207–208, 1988).

# After appropriate diet and catharsis.

** For bite wounds, in which likely pathogens may also include oral anaerobes, *Eikenella corrodens* (human), and *Pasteurella multocida* (dog and cat), some *Medical Letter* consultants recommend use of amoxicillin-clavulanic acid (Augmentin) or ampicillin-aulbactam (Unssyn).

*Adapted* from ref. 11.

## Complications of Prophylaxis

Clinical studies have generally demonstrated a paucity of side effects of prophylaxis therapy. This is most likely related to the widespread use of the cephalosporins, moderation in dosage, and relatively short duration of therapy.

Pseudomembranous colitis is the major reported complication (21). Hypoprothrombinemia and prolonged bleeding times represent more of a theoretical concern except in those patients requiring postoperative anticoagulation or undergoing procedures in which there is little tolerance for even minimal oozing (8).

The major concern is the emergence of bacteria resistant to antimicrobial agents. Colonization and subsequent secondary infection with resistant bacteria have been thought to be due to patient exposure to the resistant flora of the hospital environment, and transmission may be facilitated via the hands of hospital personnel. Recent studies (22) demonstrate that many patients are colonized with small numbers of resistant strains on admission to the hospital, and after exposure to antibiotics, these organisms proliferate. Either mechanism would have a profound impact on the continued and expanded role of prophylactic antimicrobial use, but at present, the dramatic improvement in surgical wound infection rates clearly outweighs this potential side effect (8).

## Summary

Prophylactic antibiotics play a significant role in the management of the modern surgical

TABLE 4

*Recommended Antibiotic Prophylaxis for the Most Commonly Performed Surgical Procedures*

| Surgical Procedure (rank)* | Recommended prophylactic regimen† |
|---|---|
| Gynecologic surgery | |
| Cesarean section (1) | Cefazolin (1 g IV) after clamping the cord and 6 and 12 hours later. Prophylaxis is not indicated in uncomplicated elective procedures. Uterine irrigation with antibiotics may be comparable to systemic therapy. If irrigating antibiotics are used, 2 g of cefoxitin in 1 liter of normal saline is effective. In patients with a beta-lactam allergy, metronidazole (500 mg IV) after the clamping cord is effective. |
| Dilation and curettage and abortion‡ (2) | Prophylactic antimicrobials are not recommended for uncomplicated dilation and curettage. In second-trimester instillation abortions, 1 g of cefazolin§ before the procedure and 1 g every 6 hours (2 doses) is effective. In patients with a beta-lactam allergy, oral metronidazole (400 mg) before the procedure and every 4 hours afterward (2 doses) is effective. For induced abortion in the first trimester in patients with a history of pelvic inflammatory disease, 2 million IU of penicillin G IM before and 3 hours after the procedure is effective.¶ |
| Hysterectomy, abdominal or vaginal (3) | Cefazolin (1 g IV) preoperatively and 6 and 12 hours later. Second- or third-generation cephalosporins have not proved to be more effective. In patients with a beta-lactam allergy, doxycycline (200 mg IV) preoperatively (1 dose) is effective in vaginal hysterectomy. |
| Bilateral occlusion of fallopian tubes (6) | Data to recommend prophylaxis are not available. |
| Repair of cystocele and rectocele (20) | Prophylactic antibiotics have not proved to be effective. |
| Orthopedic surgery | |
| Arthroplasty of joints, including replacement (4) | For major joint (hip or knee) repair, cefazolin (1 g IV) preoperatively and every 6 hours (3 doses) is recommended. A higher dose of cefazolin (2 g) should be considered in knee replacement when a tourniquet is used. Data on prophylaxis in arthroscopic surgery are not available. |
| Open reduction of fracture (9) | Cefazolin§ (1 g IV) preoperatively and every 6 hours (3 doses). Complex (open) fractures are considered contaminated and cefazolin§ therapy (1 g every 8 hours for 10 days beginning on admission) is indicated. |
| Laminectomy and spinal fusion‡ | Prophylactic antimicrobials have not proved to be beneficial. |
| Amputation of lower limb (24) | Cefoxitin (2 g IV) preoperatively and every 6 hours (4 doses). |
| Ophthalmic surgery | |
| Extraction of lens, including insertion of prosthesis (5) | There are no adequately controlled trials in ophthalmic surgery. A retrospective review of data suggests that antibiotics or topical antiseptics may be effective, but there is no consensus regarding efficacy or choice of therapy. |
| General surgery | |
| Cholecystectomy (7) | Cefazolin (2 g IV) preoperatively in "high-risk" patients—i.e., those older than 60 years, and those with previous biliary surgery, a history of acute symptoms, or jaundice. In patients with a beta-lactam allergy, gentamicin (80 mg IV) preoperatively and every 8 hours (3 doses) is effective. |
| Inguinal hernia repair (8) | Data to recommend prophylaxis are not available. |
| Colon surgery (13) | Neomycin and erythromycin base, 1 g of each orally at 1, 2, and 11 p.m. on the day before surgery. For emergency colon surgery or situations precluding preoperative oral prophylaxis, cefoxitin (2 g IV) preoperatively and every 4 hours (3 doses) is effective. In patients with a beta-lactam allergy metronidazole (500 mg IV) and gentamicin (1.7 mg/kg IV) preoperatively and every 8 hours postoperatively (3 doses) are effective. |
| Division of adhesions, laparotomy, and abdominal surgery not involving a viscus (15) | Data to recommend prophylaxis are not available. |
| Primary appendectomy (16) | Cefoxitin (2 g IV) preoperatively and every 6 hours (3 doses) in nonperforated appendixes; in perforated appendixes, therapy is continued for 3–5 days. Although combined aerobic and anaerobic coverage appears to be preferable, in patients with a beta-lactam allergy, metronidazole (500 mg IV) preoperatively is effective. With perforated appendixes, this is continued every 8 hours (in either IV or oral administration) for 3–5 days. |
| Mastectomy, total or partial (18) | Data to recommend prophylaxis are not available. |
| Gastric resection (26) | Cefazolin (1 g) preoperatively in "high-risk" patients only—those with a bleeding gastric or duodenal ulcer, an obstructive duodenal ulcer, a gastric ulcer, gastric cancer, or morbid obesity (prophylaxis is not indicated for cases of chronic uncomplicated duodenal ulcers). In patients with a beta-lactam allergy, 1 preoperative IV dose of gentamicin (120 mg) and clindamycin (600 mg) may be effective, but data are limited. |

TABLE 4

*Recommended Antibiotic Prophylaxis for the Most Commonly Performed Surgical Procedures (continued)*

| Surgical Procedure (rank)* | Recommended prophylactic regimen† |
|---|---|
| Surgery for penetrating abdominal trauma‡ | Cefoxitin (2 g IV) upon admission to the hospital. For patients with an intestinal perforation, 2 g of IV cefoxitin every 6 hours for 2–5 days is effective. |
| Urologic surgery | |
| Prostatectomy, transurethral and peritoneal (10) | Antimicrobial prophylaxis is not recommended in patients with sterile preoperative urine cultures. Prophylactic antibiotics have not proved to be effective in transperitoneal needle biopsies of the prostate when the preoperative cultures are sterile. |
| Dilation of urethra (21) | Data to recommend prophylaxis are not available. |
| Surgery on nose, mouth, and pharynx | |
| Tonsillectomy with or without adenoidectomy (11) | Data to recommend prophylaxis are not available. |
| Major head, neck, and oral surgery (12) | In major surgical procedures involving an incision through oral or pharyngeal mucosa, a combination of gentamicin (1.7 mg/kg) and clindamycin (300 mg IV) preoperatively and every 8 hours (2 doses) is recommended. Cefazolin and third-generation cephalosporins have also demonstrated effectiveness when given over a 24-hour period perioperatively. |
| Rhinoplasty and repair of nose (17) | Prophylactic antimicrobials have not proved to be effective. |
| Cardiothoracic and vascular surgery | |
| Median sternotomy, coronary-artery bypass grafting, and valve surgery (14) | Cefazolin§ (1 g IV) preoperatively and every 6 hours for 48 hours. In patients with a beta-lactam allergy, vancomycin (15 mg/kg IV) preoperatively, after initiation of bypass (10 mg/kg), and every 8 hours postoperatively for 48 hours may be effective. |
| Pacemaker insertion (19) | Cefazolin§ (1 g) as above; in patients with a beta-lactam allergy, no prophylaxis may be a reasonable alternative, given the low incidence of infection. |
| Thoracic surgery procedure including lobectomy and pneumonectomy (22) | Cefazolin§ (1 g) preoperatively and every 6 hours postoperatively for 24 hours. The optimal duration of postoperative prophylaxis has not been established. In penetrating thoracic trauma and in the placement of chest tubes in trauma management, prophylactic antibiotics have not been effective. |
| Peripheral vascular surgery (23) | Cefazolin (1 g) preoperatively and every 6 hours postoperatively for 24 hours. The usefulness of antibiotic prophylaxis in carotid-artery surgery has not been established, but when infection rates are high, cefazolin should be used as described above. |
| Neurosurgical procedures** | |
| Cerebrospinal fluid shunting procedures (18) | Antibiotic prophylaxis is not indicated in institutions with low infection rates (<10%). Trimethoprim (160 mg) plus sulfamethoxazole (800 mg IV) preoperatively and every 12 hours (3 doses) may be beneficial in institutions with high infection rates (>20%). |
| Craniotomy (25) | For high-risk procedures (e.g., reexploration and microsurgery), clindamycin (300 mg IV) preoperatively and at 4 hours has been found to be effective, as has vancomycin (1 g IV) plus gentamicin (80 mg IM) preoperatively. |
| Miscellaneous | |
| Simple hand laceration | Prophylactic antibiotics have not proved to be effective. However, in animal (cat) bites, oral ampicillin‖ (500 mg) every 6 hours for 5 days may be effective. |

* Procedures are ranked in descending order of the frequency with which they were performed in patients discharged from nonfederal, short-stay hospitals in 1984. The number of procedures performed ranged from 812,000 for cesarean section (Rank 1) to 33,000 for gastric resection (Rank 26). These data were compiled from U.S. Public Health Service publications and from unpublished data from the National Hospital Discharge Survey, National Center for Health Statistics, Hyattsville, Md. Diagnostic biopsies, minor office procedures, minor obstetrical and gynecologic procedures, and procedures on newborn infants are not included.

† Unless otherwise noted, the preoperative dose of intravenous (IV) antibiotic should be administered at about the time of the operative incision. Preoperative intramuscular (IM) antibiotic should be administered one-half to one hour before the incision.

‡ Coding procedures did not allow a precise categorization or tabulation of procedures.

§ Cefazolin is recommended at the indicated dose schedule, even though another dose of cefazolin, another cephalosporin, or a semisynthetic penicillin was evaluated in the referenced article.

¶ In the referenced article, oral pivampicillin (350 mg three times a day for four days) was also given after the procedure to patients in the prophylaxis group. Whether oral ampicillin (500 mg three times a day) can be substituted or whether an oral antibiotic regimen is necessary at all after the procedure is not known.

‖ Ampicillin is recommended, although oxacillin was evaluated in the referenced article.

** See "Orthopedic Surgery" section for recommendations regarding laminectomy and spinal fusion.

*Adapted* from Kaiser (ref. 10).

TABLE 5

*Prospective Randomized Comparisons of Long Versus Short Courses of Perioperative Prophylaxis*

| | Duration | | | % wound infections | | |
|---|---|---|---|---|---|---|
| Surgical procedure and antibiotic regimen | Short course | Long course | Total no. of patients studied | Short course | Long course | p Value |
| Major joint repair | | | | | | |
| Nafcillin, 1 g q6h | 24 hr | 72 hr | 358 | 1.6 | 2.3 | NS |
| Cefamandole, 2 g initially, then 1 g q6–10h | Initial dose only | 24 hr | 717 | 5.2 | 0.9 | <0.001 |
| Cesarean section | | | | | | |
| Cefuroxime, 0.75 g q8h | 24 hr | 5 days | 39 | 0* | 10 | NS |
| Ampicillin, 2 g initially, then 1 g q6h | 18 hr | 72 hr | 83 | 28* | 8 | <0.05 |
| Vaginal hysterectomy | | | | | | |
| Cefazolin, 1 g q6h | Initial dose only | 18 hr | 90 | 8.9† | 4.4 | NS |
| Median sternotomy | | | | | | |
| Cephalothin, 1 g q6h | Initial dose only | 4 days | 64 | 6.6‡ | 2.9 | NS |
| Cephalothin, 2 g q6h | 48 hr | 6 days | 200 | 2.1§ | 2.8 | NS |
| Cephalothin, 12 mg/kg, plus kanamycin, 7 mg/kg, with or without oral cephalexin, 1 g q6h × 3 days | 48 hr | 48 hr + oral cephalexin | 160 | 2.5 | 1.2 | NS |

NS: not significant.
* Endometritis.
† Cuff cellulitis.
‡ Includes one case of endocarditis.
§ More urinary tract infections developed in 2-day (17.0%) versus 6-day (8.5%) patients ($p = 0.07$).
*Adapted* from Kaiser (ref. 8).

patient. Guidelines for their use have been presented. Marked variations in nosocomial pathogens and their resistance patterns exist among institutions and surgical procedures are frequently far from routine owing to a variety of perioperative circumstances. A willingness to monitor and alter protocols based on local observations and newly published data remains imperative.

### References

1. Cruse P. Wound infection surveilance. Rev Infect Dis 1981; 3:734–737.
2. Ad Hoc Committee of the Committee on Trauma. National Research Council Division of Medical Sciences. Influence of ultraviolet irradiation of the operating room and of various other factors Ann Surg 1964; 160(2).
3. Olson M, O'Connor M, Schwartz M. Surgical wound infections; a 5 year prospective study of 20,193 wounds at the Minneapolis V.A. Medical Center. Ann Surg 1984; 199:253.
4. Culbertson WR, Altemeier WA, Gonzalez LL, et al. Studies on the epidemiology of post-operative infection of clean wounds. Ann Surg 1961; 154:599–610.
5. Howe CW, Marston AT. A study on sources of post-operative staphylococcal infection. Surg Gynecol Obstet 1962; 115:266–275.
6. Burke JF. Identification of the sources of staphylococci contaminating the surgical wound during operation. Ann Surg 1963; 158:898–904.
7. Burke JF. The effective period of preventive antibiotic action in experimental incisions and dermal lesions. Surgery 1961; 50:161–168.
8. Kaiser AM. Post-operative infections and antimicrobial prophylaxis. In: Mandel G, Douglas G, Bennett J, eds. Principles and practice of infectious disease, 3rd ed. New York: Churchill Livingstone, 1989.
9. Guglielmo BJ, Hohn DC, Koo PJ, Hunt TK, Sweet RL, Conte JE Jr. Antibiotic prophylaxis in surgical procedures: a critical analysis of the literature. Arch Surg 1983; 118:943–955.
10. Kaiser AB. Antimicrobial prophylaxis in surgery. N Engl J Med 1986; 315:1129–1138.
11. Antimicrobial prophylaxis in surgery. Med Lett 1989; 31:105–108.
12. Evans M, Pollack AV. Trials on trials: a review of trials of antibiotic prophylaxis. Arch Surg 1984; 119:109–113.
13. Freiman JA, Chalmers TC, Smith H Jr, Kuebler RR. The importance of type II error and sample size in the design and interpretation of the randomized control trial: survey of 71 "negative" trials. N Engl J Med 1978; 299:690–694.
14. Kaiser AB. Overview of cephalosporin prophylaxis. Am J Surg 1988; 155(Suppl 5A):52–55.
15. Slama TG, Sklar SJ, Misinski J, et al. Randomized comparison of cefmandole, cefazolin and cefuroxime prophylaxis in open-heart surgery. Antimicrob Agents Chemother 1986; 29:744–747.

16. Durack DT, Iswo AL, Kaplan EL. Apparent failures of endocarditis prophylaxis: analysis of 52 cases submitted to a national registry. JAMA 1983; 250:2318.
17. Schidman ST, et al. Committee on Rheumatic Fever and Infectious Carditis of the Council on Cardiovascular Disease in the Young. Circulation 1984; 70:1123A.
18. Med Lett 1987 (12/4); 29:107.
19. Kaiser AB. Effective and creative surveillance and reporting of surgical wound infections. Infect Control 1982; 3:41–43.
20. Goldmann DA, Hopkins CC, Karchner AW. Cephalothin prophylaxis in cardiac valve surgery: a prospective, double blind comparison of two-day and six-day regimen. J Thorac Cardiovasc Surg 1977;470–479.
21. Bohnen JM, Matlow AG, Cohen MM. Pseudomembranous colitis and wound infection following perioperative use of multiple antibiotics. Can J Surg 1985; 28:432–433.
22. Kernodle DS, Barg NL, Kaiser AB. Low-level colonization of hospitalized patients with methculim-resistant coagulase-negative staphylococci and their emergence during antimicrobial prophylaxis. Antimicrobial Agents Chemother 1988; 32:202–208.

# Prevention of Deep Venous Thrombosis and Pulmonary Embolism in Surgical Patients

JONATHAN L. HALPERIN, M.D.

UNDERSTANDING the pathophysiologic processes involved in deep venous thrombosis and pulmonary embolism is essential to the successful prevention of thromboembolic complications in patients undergoing surgical procedures.

**History.** The triad of stasis, endothelial injury, and a hypercoagulable state, first postulated by Rudolph Virchow over a century ago, still forms the basis of our notions of causation and prevention (1). Over the past 50 years effective clinical strategies have evolved because of refined theories of the mechanisms of physiological coagulation and pathologic thrombosis, improved tests for measuring these two distinct processes, the development of antithrombotic therapy, and the growing number of clinical trials that support specific interventions for designated populations at risk.

Relatively crude theoretical knowledge of coagulation and thrombosis prevailed at the beginning of this era, even when prothrombin time and clotting time tests were already in use. Heparin and dicumarol have been commercially available to two generations of physicians. The limited success of venography when it was originally introduced was followed by rapid advances in blood-clotting theories and the introduction of warfarin and the activated partial thromboplastin time test, as well as the publication of the landmark study by Barritt and Jordan (2) demonstrating the efficacy of heparin and oral anticoagulants for prevention of recurrent venous thrombosis and pulmonary embolism. In 1958, though, Wessler and Deykin pointed out the deficiencies imposed by the lack of an accurate diagnostic test for thrombosis (3).

Over the next 10 years, advances in the understanding of the molecular basis of thrombus formation led to the development of two mutually exclusive regimens of heparin therapy: a low-dose subcutaneous approach, given prophylactically to patients at risk of thrombosis, and full-dose heparin followed by oral anticoagulants, after frank thrombosis had occurred. More definitive clinical diagnostic tests also became available during this period, including pulmonary angiography, radioisotopic lung scanning, radioiodinated fibrinogen uptake scanning, impedance plethysmography, and Doppler ultrasound.

Over the last 10 years, our concepts of variable intensity anticoagulation have matured with the introduction of the international normalized ratio (INR), which allows comparison of prothrombin time data obtained with different thromboplastin preparations (4). The INR is calculated by raising the prothrombin time ratio (patient:control) exponentially to the power of the international sensitivity index (ISI) defined for each thromboplastin reagent in comparison with an international reference standard, the Manchester comparative reagent (which is assigned a value of 1). For tests performed with conventional North American rabbit brain thromboplastin reagents, which have ISI values between 2.0 and 2.6, low-intensity anticoagulation (PT ratio 1.5–2.0 times control) corresponds to an INR between 3.0 and 4.8, whereas low-intensity anticoagulation (PT ratio 1.3–1.5 times control) is produced at an INR between 1.7 and 3.5.

In the International MultiCentre Trial (5), one of the first large prospective prevention trials

---

From the Division of Cardiology, Mount Sinai School of Medicine, New York, NY 10029. Address reprint requests to the author at Box 1030, Division of Cardiology, Mount Sinai Medical Center, One Gustave L. Levy Place, New York, NY 10029.

involving general surgical patients, those given subcutaneous low-dose heparin had significantly fewer deaths attributed to pulmonary embolism than those given placebo. Prophylactic low-dose heparin reduced the incidence of deep venous thrombosis and prevented the extension of calf vein thrombi into the proximal veins of the thigh and groin, which leads to pulmonary emboli. Only a few years ago, all data on pulmonary embolism and mortality available from evenly randomized trials of perioperative subcutaneous heparin in general, orthopedic, and urologic surgery were pooled by Collins et al. (6) for the purpose of meta-analysis. In all, 6,366 patients were assigned to receive heparin, and there were 6,426 in the control groups. The incidence of pulmonary embolism was reduced from about 3% to 1% with heparin therapy.

In the present decade, there persists a need for more detailed fundamental understanding of the pathogenesis of deep venous disease, and for the development of standardized tests addressing the antithrombotic rather than the anticoagulant actions of heparin and warfarin. Better noninvasive tests for venous disease are sought, as are trials to determine the optimum intensity of warfarin therapy. It is to be hoped that examination of the proper role of thrombolytic therapy will be forthcoming, and perhaps the development of new antithrombotic agents with better therapeutic ratios of antithrombotic effect to bleeding risk.

**Morbidity and Mortality.** Deep venous thrombosis is responsible for the hospitalization of over half a million Americans annually and is one of the most frequent complications of other illnesses and surgical procedures. The outcome of deep venous thrombosis is too often pulmonary thromboembolism, which accounts for well over 50,000 mortalities each year in the United States, and survivors may suffer the consequences of chronic venous insufficiency, which affects nearly 20% of the adult population.

At least 90% of pulmonary emboli arise from thrombi in the deep veins of the lower extremities, though symptoms and physical signs of deep venous thrombosis may be minimal or absent. Although death may occur quickly following acute pulmonary embolism—often within 30 min—most of these episodes could be prevented through judicious use of either mechanical methods directed at reduction of venous stasis or administration of medication designed to inhibit thrombosis. The effectiveness of the former is limited, whereas the latter carries the risk of hemorrhage, so a prudent approach takes an assessment of thromboembolic risk into account as prophylactic therapy is selected from among a wide range of options.

### Assessment of Thromboembolic Risk

Clinical features predisposing to the development of deep venous thrombosis in the surgical patient include

- age over 40 years
- previous venous thromboembolic disease
- varicose veins or other forms of chronic venous insufficiency
- malignancy
- congestive heart failure
- obesity
- prolonged immobilization
- paralysis
- estrogen or corticosteroid therapy
- type and duration of surgical procedure.

**Low Risk.** A stratification of risk has been proposed (7) whereby low-risk patients are those under age 40 years without other risk factors who are undergoing surgery under general anesthesia lasting less than 30 min or elective abdominal or thoracic surgical procedures. Even if no prophylactic measures are provided, these patients have a risk of calf venous thrombosis under 10%, less than 1% chance of proximal deep venous thrombosis, and no more than 0.01% danger of fatal pulmonary embolism.

**Moderate Risk.** Patients at moderate risk are typically over age 40 years, undergoing surgery with general anesthesia lasting longer than 30 min, with one or more of the risk factors delineated above. They carry a 10%–40% risk of developing calf vein thrombi, 2%–10% risk of proximal thrombosis, and between 0.1% and 0.7% risk of fatal embolism.

**High Risk.** Those at highest risk are older patients undergoing lengthy procedures under general anesthesia who have a previous history of venous disease and patients having abdominal surgery for malignant disease or orthopedic procedures of the lower extremities. In such cases the risk of calf vein thrombosis may be as high as 80%, of proximal venous thrombi between 10% and 20%, and of fatal pulmonary thromboembolism up to 5% if no prophylactic measures are taken.

### Methods of Reducing Risk

**Low Risk Patients: Mechanical Compression.** Mechanical compression of the lower extremities is designed to reduce the venous stasis that leads to thrombus formation during and after major or minor surgery. Among the simplest mechanical measures are elastic support stockings, which provide about 20 mmHg external compression at the level of the ankle, with graded reduction in counterpressure as the garment ascends toward the knee or inguinal region. This helps minimize venous valvular incompetence and promote venous emptying.

Application of gentle heat to the limbs accelerates venous flow velocity, and this may help reduce stasis which contributes to thrombus formation.

External pneumatic compression sleeves and related intermittent pneumatic compression devices propel blood centrally from the limb circulation, more closely mimicking the physiological musculovenous pumping action associated with ambulatory activity.

Each of these methods seems to offer some benefit in terms of reducing the risk of deep venous thrombosis complicating abdominal surgical procedures without contributing any hemorrhagic risk. The prophylactic efficacy of mechanical measures and antithrombotic medication appears additive; the use of mechanical compression alone, without medication, may not be sufficient to protect many patients other than those at low risk of thromboembolism related to venous disease.

**Moderate Risk Patients: Low-Dose Heparin in Early Intervention.** The regimen of administering heparin, 5,000 units subcutaneously every 12 hr, has been shown to be effective in reducing the incidence of deep venous thrombosis in moderate-risk patients undergoing nonorthopedic surgical procedures, as well as in patients at bed rest following acute myocardial infarction and stroke, even without consistent prolongation of partial thromboplastin time.

The mechanism of action of heparin is suspected to involve prevention of activation of factor Xa in the coagulation cascade, and effectiveness seems to depend on early prophylactic intervention (8–11). For surgical patients, heparin is begun 2 hr preoperatively and continued postoperatively until the patient is discharged or becomes ambulatory. Although minor bleeding in the form of ecchymoses at the injection sites and wound hematomas have been reported in some studies, the incidence of major hemorrhage is no greater in patients given heparin in this dosage perioperatively than in saline-treated controls.

In an international multicenter trial (5) involving 4,121 patients over age 40 years undergoing elective major surgery, low-dose subcutaneous heparin (5,000 units every 12 hr) prevented both nonfatal and fatal pulmonary emboli. The incidence of clinical deep venous thrombosis extending proximal to the knee was 6% among patients in the control group, and this was reduced to 0.6% among those given heparin prophylaxis. There were 180 deaths in the in-hospital postoperative period (mortality rate 4.37%). Of 100 deaths in the control group, postmortem examinations were obtained in 72% and pulmonary embolism was documented in 16 cases; there were 80 deaths among heparin-treated patients, with a 66% autopsy rate, and only 2 instances of pulmonary embolism were confirmed.

**Failures of Low-Dose Heparin.** Initiation of heparin after surgery or when preoperative immobilization has been prolonged without prophylaxis may reduce the benefit of this approach, and the regimen has not been found effective in orthopedic surgical procedures involving the lower extremities, particularly surgery for repair of hip fracture or elective total hip replacement. The reason for the failure of low-dose heparin prophylaxis in these situations may be related to earlier activation of the coagulation system leading to fibrin formation, or to greater exposure of surface collagen at the level of damaged venous endothelium. The latter might entail activation of platelet aggregation, which is not inhibited by heparin or other anticoagulant drugs.

**Heparin-Dihydroergotamine Combination.** Heparin has been combined with the vasoconstrictor compound dihydroergotamine in a ratio of 10,000:1 in an effort to enhance its prophylactic value (12). This substance appears to prolong the kinetics of heparin action after subcutaneous injection, to promote prostaglandin synthesis, and to accelerate fibrinolysis, each of which could intensify the antithrombotic effect, but such products are not presently approved for use in the United States (13).

**Adjusted-Dose Heparin.** Adjusted-dose heparin based on the intensity of anticoagulant effect on the partial thromboplastin time appeared to reduce the incidence of deep venous thrombosis in hip replacement surgery in one study of 99 patients (14). In comparison with the fixed-dose regimen, heparin dose was adjusted upward when the kaolin-activated partial thromboplastin time obtained 6 hr after injection was less than 31.5 sec, and the dose was reduced when the aPTT was greater than 36 sec. The incidence of deep venous thrombosis in patients receiving the fixed dosage was 39%, and in those given the adjusted dosage 13%, but confirmatory data are lacking.

**Heparinoid Congeners.** A promising approach is subcutaneous administration of low-molecular-weight heparinoids, which selectively increase inhibition of factor Xa. In comparison with conventional heparin, several of the low-mo lecular-weight heparin compounds seem superior for prevention of deep venous thrombosis following abdominal surgery and elective total hip replacement (15, 16). It has also been suggested that these novel heparinoids may carry a lower risk of heparin-associated thrombocytopenia, adding to their therapeutic advantage. No low-molecular-weight heparin products are yet approved for marketing in the United States, and price relative to conventional sodium or calcium heparin salts remains a concern.

**High-Risk Patients Undergoing Orthopedic Surgery.** The problem of prophylaxis against deep venous thrombosis in high-risk cases becomes particularly challenging with orthopedic surgical procedures involving the hip and femur. Here the incidence of deep venous thrombosis is increased by the extent of perioperative immobilization and by the direct vascular trauma associated with these procedures. When acute fractures are involved, there is also a comparatively high rate of traumatic bleeding complications which make antithrombotic therapy more difficult.

There is also the problem of accurate diagnosis: the cardinal physical signs of acute thrombophlebitis—inflammatory erythema, warmth and tenderness, asymmetrical limb swelling, venous distention, and palpable cord-like subcutaneous masses—are frequently encountered even in uncomplicated postoperative limbs of orthopedic patients; noninvasive tests such as impedance plethysmography and Doppler velocity analysis have low sensitivity and specificity in this setting, and the postoperative hip surgery patient poses complex challenges even to sophisticated laboratories offering such methods as phleborheography, indium 111 oxine platelet scintigraphy, or duplex ultrasound vascular imaging, so contrast phlebography has been accepted as the standard diagnostic technique. This makes the assessment of prophylactic or therapeutic outcome cumbersome, and data from clinical trials have as yet provided no clear direction on which to base recommendations.

**Heparin Regimens after Orthopedic Surgery.** The approach to prevention of deep venous thrombosis following orthopedic surgery remains difficult and controversial. Use of subcutaneous heparin in the low dose of 5,000 units every 12 hr has not generally offered sufficient protective benefit, and full anticoagulant doses, administered either intermittently or by continuous intravenous infusion, are associated with increased risk of bleeding. Various adjusted-dose regimens have been proposed, directed at raising the partial thromboplastin time minimally above that of normal control plasma, but no single approach has been validated in properly designed, placebo-controlled trials. Data from a multicenter study comparing low-molecular-weight heparin and nonfractionated heparin in the prevention of deep vein thrombosis and pulmonary embolism do not reveal significant differences (16), nor do results differ when low-molecular-weight heparin is compared to fractionated heparins as analyzed by metaanalysis (16A).

**Warfarin after Orthopedic Surgery.** Oral or intramuscular warfarin has been advocated as an alternative to heparin in patients undergoing orthopedic surgical procedures involving the lower extremities (17). The anticoagulant is initiated preoperatively and the dose is adjusted postoperatively to prolong the prothrombin time to 1.12–1.2 times the control value, corresponding to an international normalized ratio (INR) of prothrombin suppression in the range of 1.2–1.5. Recent studies suggest that the prothrombin time is a biologically insensitive assay of antithrombotic effect, and other tests, such as factor VIIc activity or the F1 + 2 fragments of prothrombin, may detect changes in the coagulation properties of blood with doses of warfarin insufficient to perturb the prothrombin time. One study, in patients undergoing gynecologic surgery, found small doses of oral warfarin, in the range of 1 mg daily, effective in comparison with placebo in prevention of postoperative deep venous thrombosis (18). Similar effectiveness has been reported in maintaining the patency of indwelling central venous catheters prone to thrombotic occlusion in patients undergoing treatment of malignancies (19).

These results raise the possibility that low doses of anticoagulant medication, which are not sufficient to produce measurable changes in standard tests of coagulation, may offer some prophylactic value against intravascular thrombosis.

**Dextran-70.** Dextran-70 infusions are thought to exert some antithrombotic effect mainly through hemodilution and increased venous flow velocity, but objective evidence of benefit is less impressive than with anticoagulant drugs (20–24). A potential advantage is a lower risk of hemorrhagic complications, but one disadvantage is the possibility of fluid overload, with adverse cardiac and renal consequences.

**Platelet-Inhibiting Agents.** Aspirin as a platelet inhibitory agent has been examined as a clinical antithrombotic agent in a number of settings involving arterial disease and other cardiovascular conditions, with wide-ranging success (25–33). Consistent demonstration of effectiveness in interference with venous thrombosis has not been forthcoming, as stasis with formation of thrombin and fibrin is thought to predominate pathophysiologically over vascular wall endothelial injury with platelet aggregation. There may still be a role for inhibiting platelet activation when surgical procedures involve retraction or manipulation of large veins, as in orthopedic procedures involving the hip or venous procedures themselves. Trials of heparin in combination with platelet inhibitor drugs, such as aspirin, dipyridamole, sulfinpyrazone, or the investigational agent ticlopidine, have not been completed.

**Cost Effectiveness.** Available methods of prophylaxis against postoperative deep venous thrombosis and pulmonary embolism have been cost-effective for high-risk patients (34).

**Potential Improvements.** On the horizon are even more accurate noninvasive diagnostic methods and improved clinical instruments for stratification of thromboembolic risk, from which more appropriate application of these interventions can be expected. Understanding of the interactive biology of the blood and vessel wall is yielding tests of prothrombotic and antithrombotic activity superior to those presently employed, and this should translate into more effective clinical strategies. New drugs, such as specific thrombin inhibitors, hold the promise of higher therapeutic ratios separating prophylactic benefit from the potentiation of bleeding. The clinician will be called on increasingly to draw from a broad array of therapeutic agents which must be applied in an individualized fashion appropriate to the specific circumstances of each case.

## References

1. Virchow R. Neuer fall von todlichen: emboli der lungenarterie. Arch Pathol Anat 1856; 10:225–228.
2. Barritt DW, Jordan SC. Anticoagulant drugs in the treatment of pulmonary embolism: a controlled clinical trial. Lancet 1960; 1:1309–1312.
3. Wessler S, Deykin D. Theory and practice in acute venous thrombosis—a reappraisal. Circulation 1958; 18: 1190–1214.
4. Hirsh J, Poller L, Deykin D, Levine M, Dalen JE. Optimal therapeutic range for oral anticoagulants. Chest 1989; 95(Suppl):5S–11S.
5. The International MultiCentre Trial. Prevention of fatal postoperative pulmonary embolism by low doses of heparin. Lancet 1975; 2:45–51.
6. Collins R, Scrimgeour A, Yusuf S, Peto R. Reduction in fatal pulmonary embolism and venous thrombosis by perioperative administration of subcutaneous heparin: overview of results of randomized trials in general, orthopedic, and urologic surgery. N Engl J Med 1988; 318:1162–1173.
7. Merli GJ, Martinez J. Prophylaxis for deep vein thrombosis and pulmonary embolism in the surgical patient. Med Clin North Am 1987; 71:377–397.
8. Kelton JG, Hirsh J. Bleeding associated with antithrombotic therapy. Semin Hematol 1980; 17:259–291.
9. Kiil J, Kiil J, Axelson F, et al. Prophylaxis against postoperative pulmonary embolism and deep venous thrombosis by low-dose heparin. Lancet 1978; 1:115–116.
10. String ST, Barcia PJ. Complications of small dose prophylactic heparinization. Am J Surg 1986; 130:570–574.
11. vanOoijen B. Subcutaneous heparin and postoperative wound hematomas. Arch Surg 1986; 121:937–940.
12. Lahnborg G. Effect of low dose heparin and DHE prophylaxis on frequency of postoperative deep vein thrombosis in patients undergoing post-traumatic hip surgery. Acta Chir Scand 1980; 146:319.
13. Multi-Center Trial Committee. DHE/heparin prophylaxis of postoperative DTV. JAMA 1984; 251:22, 2960–2966.
14. Leyvraz P, Richard J, Bachman F, et al. Adjusted versus fixed dose subcutaneous heparin in the prevention of DVT after total hip replacement. N Engl J Med 1983; 309:954.
15. Turpie AG, Levine MN, Hirsh J, et al. A randomized controlled trial of a low molecular weight heparin to prevent DVT in patients undergoing elective hip surgery. N Engl J Med 1986; 15:925–929.
16. Kakkar VV, Stringer MD, Hedges AR, et al. Fixed combinations of low-molecular-weight or unfractionated heparin plus dihydroergotamine in the prevention of postoperative deep vein thrombosis. Am J Surg 1989; 157:413–418.

16A. Daures JP, Schved JF, Momas I, Gril JC, Azonlay P, Gromy F. Meta-analysis on randomized trials comparing the results of fractional heparins in the prevention of deep venous thrombosis. Rev Epidemiol Sante Publique 1989; 37(4):363–369.

17. Morris GK, Mitchell JR. Warfarin sodium in the prevention of deep venous thrombosis and pulmonary embolism in patients with fractured neck of the femur. Lancet 1976; 2:869–872.
18. Poller L, McKernan A, Thomson JM, Elstein M, Hirsch PJ, Jones JB. Fixed mini-dose warfarin: a new approach to the prophylaxis of venous thrombosis after major surgery. Br Med J 1987; 295:1309–1312.
19. Bern MM, Bothe A, Bistrian B, et al. Prophylaxis against central vein thrombosis with low dose warfarin. Surgery 1986; 99:216–220.
20. Kline A, Hughes LE, Campbell H, et al. Dextran 70 in prophylaxis of thromboembolic disease after surgery: a clinically oriented randomized double-blind trial. Br Med J 1975; 2:109–112.
21. Gruber UF, Saldeen T, Brokop T, et al. Incidence of fatal postoperative pulmonary embolism after prophylaxis with Dextran-70 and low dose heparin: an international multicentre trial. Br Med J 1980; 280:69–78.
22. Ahlberg A, Nylander G, Robertson B, et al. Dextran in prophylaxis of thrombosis in fractures of the hip. Acta Chir Scand 1968; 387(Suppl):83–85.
23. Johnson SR, Bygderman S, Eliasson R. Effect of dextran on postoperative thrombosis. Acta Chir Scand 1968; 387(Suppl):80–82.
24. Gruber UF: Prevention of fatal postoperative pulmonary embolism by heparin dihydroergotamine or Dextran 70. Br J Surg 1982; 69:54–58.
25. Dechavanne M, Ville D, Viala JJ, et al. Controlled trial of platelet antiaggregating agents and subcutaneous heparin in prevention of postoperative deep vein thrombosis in high risk patients. Haemostasis 1975; 4:94–100.
26. Guyer RD, Booth RE Jr, Rothman RH. The detection and

prevention of pulmonary embolism in total hip replacement: a study comparing aspirin and low dose warfarin. J Bone Joint Surg (Am) 1982; 64:1040.

27. Loew D, Brucke B, Simma W. Acetylsalicylic acid, low-dose heparin, and a combination of the substances in the prevention of post-operative thromboembolism: a double-blind study. Thromb Res 1977; 1:81–86.
28. Medical Research Council. Report of the Steering Committee: effect of aspirin on postoperative venous thrombosis. Lancet 1972; 2:441–444.
29. Morris GK, Mitchell JRA. Preventing venous thromboembolism in elderly patients with hip fractures: studies of low dose heparin, dipyridamole, aspirin and flurbiprofen. Br Med J 1977; 1:535–537.
30. Schondorf TH, Hey D. Combined administration of low-dose heparin and aspirin as prophylaxis of deep vein thrombosis after hip surgery. Haemostasis 1976; 5:250–257.
31. Soreff J, Johnsson H, Diener L, et al. Acetylsalicylic acid in a trial to diminish thromboembolic complications after elective hip surgery. Acta Orthop Scand 1975; 46:246–255.
32. Harris WH, Salzman EW, Athanasoulis C, et al. Aspirin prophylaxis of venous thromboembolism after total hip replacement. N Engl J Med 1977; 297:1246–1249.
33. Harris WH, Salzman EW, Athanasoulis C, et al. Comparison of warfarin, low molecular weight dextran, aspirin and subcutaneous heparin in prevention of venous thromboembolism after total hip replacement. J Bone Joint Surg (Am) 1974; 56:1552–1562.
34. Hull RD, Hirsh J, Sackett DL, Stoddart GL. Cost effectiveness of primary and secondary prevention of fatal pulmonary embolism in high-risk surgical patients. Can Med Assoc J 1982; 127:990–995.

# The Art and Science of Consultation in Medicine

## Preoperative Evaluation

# Preoperative Evaluation of Pulmonary Status

JONATHAN T. KLEIN, M.D., AND MARK J. ROSEN, M.D.

DESPITE advances in surgical, anesthetic, and respiratory therapy techniques, respiratory complications are quite common in patients undergoing major surgical procedures; estimates of incidence range from 10% to 80%. These complications include atelectasis, respiratory center depression due to anesthetics and analgesics, bronchospasm, and infectious sequelae ranging from exacerbations of chronic bronchitis to pneumonia and respiratory failure. Since complications occur most frequently in individuals with underlying lung diseases, preoperative assessment of a patient's pulmonary function and risk of complications is essential. Based on this evaluation, measures may be employed prior to surgery to lessen these risks. The preoperative assessment may also influence intraoperative and postoperative management. If the preoperative evaluation suggests a high operative or postoperative risk, the procedure can be canceled.

## Risk Factors

### Characteristics of the Patient

A number of patient characteristics have been associated with postoperative pulmonary complications.

**Underlying Lung Disease.** Underlying chronic respiratory diseases, such as asthma, chronic bronchitis, emphysema, and restrictive pulmonary disorders, place the patient at substantial risk of postoperative complications (1–3). The majority of these patients are easily identified by a complete history, physical examination, and chest X-ray.

**Cigarette Smoking.** Cigarette smoking increases perioperative morbidity and mortality, regardless of whether the patient is symptomatic (4, 5). Even asymptomatic smokers have alterations in lung volume that predispose to pulmonary complications.

The most common abnormality is an increased closing capacity, the volume at which dependent airways undergo compression and closure. Normally, closing capacity is less than the functional residual capacity (FRC), the gas remaining in the lungs at the end of a normal tidal exhalation, and all airways remain open throughout the respiratory cycle. However, if the closing capacity exceeds the FRC, as occurs in many smokers, airway closure occurs during tidal breathing. The problem becomes worse in the postoperative period, when the FRC decreases because the patient is recumbent without a concomitant fall in closing capacity. The increased discrepancy between FRC and closing capacity causes a further decline in gas exchange.

**Obesity.** Obese patients also are at increased risk because of reduction in tidal volume and FRC associated with elevation of the diaphragm (6). This problem is worse when such patients assume the supine position, rendering them prone to early airway closure, hypoxemia, and atelectasis.

**Advanced Age.** Elderly patients may have impaired cough reflexes, predisposing them to retention of secretions and atelectasis (7, 8). Furthermore, closing capacity increases with age, causing a ventilation-perfusion mismatch similar to that seen in smokers (9, 10).

---

From the Division of Pulmonary and Critical Care Medicine, Mount Sinai School of Medicine (JTK), and the Division of Pulmonary and Critical Care Medicine, The Beth Israel Medical Center, New York, NY (MJR). Address reprint requests to Dr. Mark J. Rosen, Division of Pulmonary and Critical Care Medicine, Beth Israel Medical Center, First Avenue and 16th Street, New York, NY 10003.

## Characteristics of the Type of Surgery

The type of surgery to be performed must be considered when evaluating a patient's pulmonary risk, because each type of surgery carries different risks of pulmonary complications.

**Low Risk: Nonthoracic, Nonabdominal, or Lower Abdominal Surgery.** Nonthoracic, nonabdominal surgery causes no consistent postoperative changes in pulmonary function and poses virtually no risk of pulmonary complications (11, 12).

Lower abdominal surgery is associated with transient, slight changes in pulmonary function and a relatively low risk of pulmonary complications (12–15).

**High Risk: Upper Abdominal or Thoracic Surgery.** Upper abdominal surgery causes prolonged changes in respiratory patterns, including decreases in lung volume and hypoxemia; the risk of perioperative pulmonary morbidity and mortality is relatively high (12–16). These postoperative changes have been attributed, in part, to pain and recumbent position. However, when the supine position is accounted for and pain is eliminated by epidural block, significant alterations in pulmonary function persist (16, 17).

Alterations in pulmonary function subsequent to thoracotomy without resection of lung are similar to those of upper abdominal surgery, except that mild impairments persist longer (15, 18, 19). When thoracotomy is performed for lung resection, the physiologic effects of the thoracotomy are added to the effects of the volume of functional lung tissue resected.

## Characteristics of the Mode of Anesthesia

Almost all general anesthetics depress cough reflexes and minute ventilation, and many induce bronchospasm and stimulate sputum production. These effects carry over postoperatively. Spinal anesthesia is often assumed to confer less pulmonary risk than general anesthesia. However, there is no consistent evidence that demonstrates a difference in the incidence of postoperative pulmonary complications (20).

## Countering the Risks

Recognition of perioperative risk factors should prompt the application of prophylactic measures such as cessation of smoking, weight loss, and bronchodilator therapy. Smokers anticipating elective surgery should be urged to stop smoking immediately. All patients undergoing upper abdominal or thoracic surgery, regardless of risk factors, should be trained preoperatively in maneuvers that prevent the development of atelectasis, such as deep breathing or incentive spirometry. Training in the immediate postoperative period, when patients may be in pain or receiving analgesics, is usually unsuccessful.

## The Preoperative Pulmonary Assessment

**History.** A thorough history should be elicited, with special focus on risk factors and symptoms related to pulmonary diseases. These include cigarette smoking, exposure to pulmonary toxins such as asbestos or solvents, prior lung infection, and previous surgery or injury to the chest. Pulmonary symptoms, including cough, sputum production, and dyspnea, should be investigated thoroughly.

**Physical Examination.** The physical examination focuses on detecting signs of obstructive, restrictive, or vascular pulmonary disease. The bedside evaluation for obstructive lung disease includes an estimation of the ratio of the time of inspiration and expiration, which is normally 1:2. One may also estimate the time it takes to forcefully exhale the vital capacity, which is normally less than 4 sec. The presence of adventitious sounds such as wheezes and rhonchi is suggestive of obstructive lung disease; these sounds may be audible only during forced exhalation. Hyperinflation may be detected at the bedside. At residual volume (the volume of air remaining in the lungs after maximal expiration), the diaphragm is generally percussible posteriorly at the level of the bottom of the scapula and descends 4–6 in. on maximal inspiration. However, in obstructive lung disease with hyperinflation, the diaphragm is below the scapula at residual volume and fails to descend normally on inspiration to total lung capacity. An increase in circumference of the thoracic cage also suggests hyperinflation.

The physical examination of patients with restrictive lung disease includes inspection of the chest wall for skeletal abnormalities, such as kyphoscoliosis or ankylosing spondylitis, and evaluation of position and mobility of the diaphragm.

Vascular lung disease leading to pulmonary hypertension may be secondary to intrinsic lung or cardiac disease, or it may be primary, in the form of primary pulmonary hypertension or pulmonary embolism. Signs of pulmonary hypertension, such as right ventricular heave, loud pulmonic valve closing sound with narrowed splitting of $S_2$, pulmonic ejection sound, jugular venous distention, or peripheral edema, are evident in advanced pulmonary hypertension.

**Chest Roentgenogram.** If the history, physical examination, and chest roentgenogram do not suggest perioperative pulmonary risk, then no further evaluation is required and the planned procedure may be performed with little risk.

**Other Studies.** If a risk factor is identified, further studies of pulmonary function may be undertaken prior to surgery in order to identify individuals likely to experience complications. However, a recent review of preoperative pulmonary function testing has stressed that with the exception of patients undergoing lung resection, it is not clear that these tests are superior to a careful history and physical examination in identifying patients at risk (21).

There is a correlation between poor spirometric values and postoperative pulmonary complications (22–25). The most reliable predictors of outcome are the $FEV_1$ (the volume of air expired in the first second of forced exhalation starting from TLC) and MVV (maximum voluntary ventilation). $FEV_1$ less than 1 L, or an MVV less than 50% of the predicted value, portends a particularly high perioperative pulmonary risk. A preoperative arterial blood gas analysis that reveals hypercapnia also identifies high-risk patients, particularly in lung resection candidates (1, 26).

When lung resection is contemplated, nuclear ventilation and perfusion scans may be useful in predicting postoperative pulmonary function (27). Four patterns have been described:

- the section of lung to be resected is neither ventilated nor perfused; resection should therefore have minimal effect on overall pulmonary function
- the section of lung to be removed has normal ventilation and perfusion, and loss of function probably will be proportionate to the volume of lung removed
- the section of lung to be removed has intact ventilation and perfusion, but the lung tissue not resected is poorly functional; removal of this portion of lung tissue is likely to cause significant loss of pulmonary function
- the section of lung to be resected has normal perfusion but no ventilation and is a source of shunting of venous blood into the arterial circulation; removal of this portion of lung is likely to improve oxygenation

In patients undergoing lung resection, a predicted postoperative $FEV_1$ value less than 0.8 L increases the risk of postoperative respiratory failure and death and has been suggested as a criterion of inoperability (28).

Preoperative spirometry and arterial blood gas analysis should be performed routinely in all patients before lung resection surgery, and in any individual with a history of cigarette smoking or dyspnea scheduled to undergo upper abdominal or thoracic surgery, including coronary artery bypass procedures (21). The usefulness of these screening tests prior to other surgical procedures is unproven.

## Summary

Before any operation on the chest or upper abdomen, all patients—regardless of the presence or absence of underlying lung disease—should be evaluated with a history and physical examination and should be trained in deep breathing or incentive spirometry.

### References

1. Tisi GM. Preoperative evaluation of pulmonary function. Am Rev Respir Dis 1979; 119:293–310.
2. Belinkoff S, Berry FD. The importance of preoperative pulmonary function evaluation. Anesth Analg 1971; 50:83–85.
3. Cahill JM. Respiratory problems in surgical patients. Am J Surg 1965; 116:362–368.
4. Latimer G, Dickman M, Clinton W, et al. Ventilatory patterns and pulmonary complications after upper abdominal surgery determined by preoperative and postoperative computerized spirometry and blood gas analysis. Am J Surg 1971; 122:622–632.
5. Morton HJV, Cumb DA. Tobacco smoking and pulmonary complications after operation. Lancet 1944; 1:368–370.
6. Putnam H, Janicek JA, Cellar CA, Wilson RD. Anesthesia in the morbidly obese. South Med J 1974; 67:1411–1417.
7. Zeffren SE, Hartford CE. Comparative mortality for various surgical operations in older versus younger age groups. J Am Geriatr Soc 1970; 20:485–489.
8. Pontopiddan H, Beecher HK. Progressive loss of protective reflexes in the airway with the advance of age. JAMA 1960; 174:2209–2213.
9. Morris JF, Koski A, Johnson ZC. Spirometric standards for healthy nonsmoking adults. Am Rev Respir Dis 1971; 103:57–67.
10. Knudson RJ, Slatin RC, Libouvitz MD, Burrows B. The maximal expiratory flow-volume curve: normal standards, variability and effects of age. Am Rev Respir Dis 1976; 113:587–600.
11. Anscombe AR, Buxton RS. Effect of abdominal operations on total lung capacity and its subdivisions. Br Med J 1958; 2:84–87.
12. Parfrey PS, Harte PJ, Quinlan JP, et al. Pulmonary function in the early postoperative period. Br J Surg 1977; 64:384–389.
13. Dureuil B, Cantineau JP, Vogel J, et al. Vital capacity and diaphragmatic function after abdominal surgery. Anesthesiology 1984; 61:A478.
14. Gracey DR, Divertie MB, Didier EP. Preoperative pulmonary preparation of patients with chronic obstructive pulmonary disease. Chest 1979; 76:123–129.

15. Ali J, Weisel RD, Layug AB, et al. Consequences of postoperative alterations in respiratory mechanics. Am J Surg 1974; 128:376–382.
16. Craig DB. Postoperative recovery of pulmonary function. Anesth Analg 1981; 60:46–52.
17. Bromage PR, Camporesi E, Chestnut D. Epidural narcotics for postoperative analgesia. Anesth Analg 1980; 59:473–480.
18. Gass GD, Olsen GN. Preoperative pulmonary function testing to predict postoperative morbidity and mortality. Chest 1986; 89:127–135.
19. Woltering EA, Flye MW, Huntley S, et al. Evaluation of bupivacine nerve blocks in the modification of pain and pulmonary function after thoracotomy. Ann Thorac Surg 1980; 30:122–127.
20. Rehder K, Sessler AD, Marsh HM. General anesthesia and the lung. Am Rev Respir Dis 1975; 112:541–563.
21. Zibrak JD, O'Donnell CR, Martin K. Indications for pulmonary function testing. Ann Intern Med 1990; 112:763–771.
22. Gaensler EA, Cugell DW, Lingren I, et al. The role of pulmonary insufficiency in mortality and invalidism following surgery for pulmonary tuberculosis. J Thorac Surg 1955; 29:163–187.
23. Mittman C. Assessment of operative risk in thoracic surgery. Am Rev Respir Dis 1961; 84:197–207.
24. Stein M, Cassara EL. Preoperative pulmonary evaluation and therapy for surgery patients. JAMA 1970; 211: 787–790.
25. Stein M, Koota GM, Simon M, et al. Pulmonary evaluation of surgical patients. JAMA 1962; 181:765–770.
26. Van Nostrand D, Kjelsberg MO, Humphrey EW. Pre-resectional evaluation of risk from pneumonectomy. Surg Gynecol Obstet 1968; 127:306–312.
27. Demeester TR, Van Heertum RL, Yarns JR, et al. Preoperative evaluation with differential pulmonary function. Ann Thorac Surg 1974; 18:61–71.
28. Olsen GN, Block AJ, Swenson EW, et al. Pulmonary function evaluation of the lung resection candidate: a prospective study. Am Rev Respir Dis 1975; 11:379–387.

# Preoperative Evaluation of Hematologic Status

CESARE CALANDRI, M.D., AND JACOB H. RAND, M.D.

## Evaluation of Anemia

TWO major problems confront the medical consultant with respect to the evaluation of a preoperative patient with anemia:

- assessment of the etiology of the anemia in relation to the patient's history and the disease that necessitates surgical intervention;
- assessment of the acceptable degree of anemia given the clinical situation of the patient, the type of surgery, and its appropriate treatment either by eliminating the cause or by blood transfusion.

### Assessment of Etiology

The different causes of anemia will not be reviewed here. However, it should be emphasized that an unexpected normocytic and normochromic anemia, even if mild or moderate (for example, hemoglobin [Hb] 9–10 g/dL), should not be dismissed as being irrelevant to the surgery. The evaluation should not be postponed until the postoperative period. For example, a patient undergoing orthopedic surgery for a fractured hip may initially have a mild anemia which may be caused by an underlying multiple myeloma. The same mild anemia could be the only obvious abnormality of underlying hypothyroidism (1) or of a pancreatic cancer. A consultant should be wary of attributing anemia to "chronic disease" without having ascertained the actual presence of a chronic inflammatory disease (2).

The etiologic diagnosis of the anemia may have dramatic effects on the planned surgery: for instance, an elective operation should be postponed if the patient is hypothyroid, or truly elective surgery might be canceled if the patient is found to have a malignancy. If an immunohemolytic anemia is suspected, note that a negative Coombs' test does not definitely rule it out (3): compatible blood can be difficult or impossible to find for a patient with immunohemolytic anemia in case of hemorrhage, not to mention the possibility of precipitating a hemolytic crisis if that same patient is exposed to hypothermia if the antibody is a cold agglutinin. In these cases the input of a hematologist is necessary.

### Assessment of Treatment

There is no question that a lower hemoglobin level is associated with higher morbidity and mortality among surgical patients (4). The question is "How low is too low?" The long-held belief that the hemoglobin level should be at least 10 g/dL to avoid complications related to the anemia are not supported by the literature. This point is particularly important in view of the recognition of transfusion-related infections and transfusion reactions.

Assuming that the anemic patient has a normal blood volume, oxygen delivery to tissues is directly related to the Hb level (1 g of Hb carries 1.34 mL oxygen), cardiac output, oxygen saturation (which can be affected by lung disease), and tissue oxygen extraction (intact microvasculature and P50 in the oxygen-Hb dissociation curve) (5). Therefore, an anemic patient with a cardiac problem, pulmonary disease, or vascular disease is more likely to experience poor peripheral oxygen delivery than a patient with normal Hb value and heart, lung, and vessels.

---

From the Hematology Division, Department of Medicine, The Mount Sinai School of Medicine, New York, NY 10029. Address reprint requests to Jacob H. Rand, M.D., Hematology Division, Department of Medicine, The Mount Sinai School of Medicine, Box 1079, One Gustave L. Levy Place, New York, NY 10029.

Even though general anesthesia and surgery carry a risk of hypoxia, certain procedures are more likely to cause hypoxia, either because of blood loss or because of the nature of the intervention. In fact, a blood loss of more than 500 mL has been associated with significantly higher incidence of complications in patients with Hb < 10 g/dL (4).

In summary, if the surgery is minor and the patient is otherwise healthy, an Hb level of 7–8 g/dL is perfectly tolerable and no blood transfusion is necessary. If the patient has severe chronic obstructive pulmonary disease, even if the surgery is minor but still requires general anesthesia, the Hb level should be raised to at least 10 g/dL. If significant blood loss is likely or if the general health status of the surgical candidate is questionable, the Hb level should be around 10 g/dL.

As far as specific treatment of the anemia is concerned, it is preferable to treat the underlying cause and to avoid blood transfusions. If surgery is emergent, and the anemia cannot otherwise be treated, then packed red blood cells should be given as a last resort. Preferably, blood should be donated by a family member or by a donor known to the patient ("directed donation"). It is good practice not to transfuse a patient, especially if he or she is elderly or has multiple medical problems, immediately prior to surgery; whenever possible, it is advisable to give blood 3 or 4 days prior to the intervention: this will minimize problems related to undetected vascular overload or early transfusion reactions (it is very difficult to diagnose and consequently treat transfusion reactions during or immediately after surgery) (6).

Autotransfusion is a relatively recent method of treating postoperative anemia (7); its role in the treatment of preoperative anemia is obviously limited. Intraoperative autotransfusion refers to the collection, filtration, and reinfusion of blood lost during surgery: this is not yet a widespread practice and has its limitations, including the risks of infections, favoring metastasis in a cancer patient, thrombocytopenia, and so on (8), but could be of use in patients who are Jehovah's Witnesses to minimize the occurrence of postoperative anemia. The availability of recombinant erythropoietin for anemic patients undergoing elective surgery may reduce the use of homologous blood in the future, not only in patients with renal failure, but also in patients with other types of chronic anemia or in patients who are Jehovah's Witnesses, especially when significant blood loss is foreseen.

A special group of patients are those affected by hemoglobinopathies; individuals with sickle cell anemia should not be routinely transfused just to raise their Hb level unless indicated on the basis of the above guidelines. If the patient is to undergo major surgery with the possibility of hypoxia and bleeding, it is advisable to reduce the Hb S level to less than 40% (9), by exchange transfusion in order to reduce viscosity.

## Evaluation of Risk of Thrombosis and Bleeding

The goal of the internist assessing the surgical candidate from the hemostatic point of view (which includes evaluation of bleeding and thrombotic risk) is to minimize the risk of thrombosis or excessive bleeding occurring during the procedure or postoperatively. The evaluation hinges on four criteria: type of surgery to be performed; whether it is an elective or an emergency procedure; history and physical examination; laboratory data, if necessary.

### Assessment of Thrombotic Risk

Patients at risk of developing deep vein thrombosis (DVT) and pulmonary embolism (PE) can be divided into three major risk groups (10):

- Low risk: age < 40 years; no other risk factors
- Moderate risk: age > 40 years; surgery lasting more than 30 min
- High risk: prior history of DVT or PE; major orthopedic surgery (particularly total hip or knee replacement); surgery for malignant disease; urologic surgery (particularly prostate)

Other factors that add to the risk of developing thrombosis are obesity, pregnancy, recent myocardial infarction or congestive heart failure, birth control pill, stroke or immobilization, varicose veins.

Prophylaxis for venous thrombosis is usually achieved with the use of heparin, typically at a dose of 5,000 units subcutaneously every 12 hr, started immediately before surgery and continued until mobilization occurs. For the patient who is to undergo orthopedic surgery, adjusted low-dose heparin (4,000 units subcutaneously every 8 hr) is recommended to maintain the activated partial thromboplastin time just above the normal laboratory range throughout the operative and postoperative period until discharge (11). No major bleeding complications have been observed with the use of these regimens. If bleeding is to be avoided, the best alternative approach is the use of intermittent external pneumatic compression of the lower extremities.

Patients with evidence of "hypercoagulability"—for example, those having a positive history of thrombosis and abnormalities such as antithrombin III, protein C, and protein S deficiencies or lupus anticoagulant/anticardiolipin antibody syndrome—may require more intense treatment. In these cases the patients have usually been maintained on warfarin, which needs to be discontinued. Anticoagulation with heparin should be continued until approximately 4–6 hr before the scheduled surgery and resumed as soon as possible (within 24–48 hr) after it (11). Other antithrombotic drugs now under investigation are likely to change the approach to venous thrombosis prophylaxis in the future; these include low-molecular-weight heparins and heparinoids, recombinant antithrombin III, and hirudin.

## Assessment of Bleeding Risk

**History and Physical Examination.** Family history is most important in relation to hereditary bleeding disorders. However, the absence of a positive family history does not rule them out. For example, 30% of the cases of classical hemophilia (factor VIII deficiency) are "sporadic," that is, due to a new mutation (12). History of previous surgery, especially if recent, is of paramount relevance in the assessment of possible bleeding (13). Dental extraction is usually associated with bleeding, sometimes profuse and lasting more than a day or resuming after it stopped, in patients with various bleeding disorders. The same is true for tonsillectomy or adenoidectomy. Excessive bleeding in relation to menses or labor and delivery warrants a full investigation. A history of absence of bleeding at circumcision or during infancy may be less significant in an adult being evaluated, as compared to a child, because of the possibility of acquired bleeding disorders during the intervening years.

Questions about the present medical or surgical history are also important: disseminated intravascular coagulation (DIC) can be associated with cancer (14) or inhibition of platelet function, and coagulation factor inhibitors can occur with monoclonal gammopathies and autoimmune diseases (15); abnormal bleeding can also be present in patients with connective tissue diseases, such as osteogenesis imperfecta, or collagen vascular diseases, such as systemic lupus erythematosus. Recurrent epistaxis and ecchymoses, especially if of recent onset, also suggest the presence of a bleeding problem.

A major point to be assessed during history taking is the use of drugs that can interfere with normal hemostasis. Although aspirin is certainly the most common culprit, bear in mind other nonsteroidal anti-inflammatory drugs or over-the-counter medications which contain salicylates or other nonsteroidal anti-inflammatory agents. Some penicillins, amitriptyline, and ethanol can also be associated with abnormal platelet function.

The physical examination can offer clues to the presence of a possible bleeding disorder. Petechiae and multiple ecchymoses usually indicate a platelet defect. The presence of organomegaly should prompt investigations regarding liver function and platelet count. The presence of deep hematomas, hemarthrosis, or joint deformities may point to underlying, previously undiagnosed hemophilia, although the majority of patients with this disorder will most likely have been diagnosed before coming to the attention of an internist.

**Laboratory Data.** The most recent reviews have cast doubt on the utility of screening coagulation studies in the absence of a significant clinical history for bleeding (13, 16, 17). In patients who have a significant positive history for previous bleeding, four tests should be performed: bleeding time (BT), prothrombin time (PT), activated partial thromboplastin time (PTT), and platelet count (PC).

**Bleeding Time.** Even though no studies have clearly correlated prolonged BT with intraoperative and postoperative bleeding (18, 19), this test is still the only available method of screening for a platelet function abnormality. If the BT is significantly prolonged in the presence of a normal platelet count, then hematologic consultation is indicated.

**Prothrombin Time and Activated Partial Thromboplastin Time.** Both PT and PTT will help in assessing the coagulation factors. A prolonged PT can indicate multiple factor deficiencies (such as in liver disease or vitamin K deficiency), but can also indicate single-factor deficiency (particularly VII, X, and fibrinogen, less commonly hypoprothrombinemia). Prolonged activated PTT can be present in liver disease and vitamin K deficiency but can also be present in hemophilias (A, B, or C), contact activation factor abnormalities, inhibitors (that is, antibodies against any of the coagulation factors or von Willebrand factor), lupus anticoagulant (again no risk of bleeding unless thrombocytopenia or antiprothrombin antibodies are also present).

The PT and activated PTT may be normal even if a coagulation factor level is significantly reduced (for instance, a factor VIII level of 30% usually corresponds to a normal activated PTT). For this reason it is important to pay attention to even minimal elevations of these tests and repeat them to verify whether an abnormality actually exists. An abnormally short activated PTT can be seen in subclinical or early overt DIC (20) and has also been associated with higher incidence of postoperative thrombosis (13).

If all tests are normal and in the face of evidence of significant history of bleeding, fibrin degradation products (DIC), platelet function tests, ristocetin cofactor (for von Willebrand disease), thrombin time (for qualitative fibrinogen abnormalities, especially in liver disease), alpha 2-antiplasmin, and a clot stability test (for factor XIII deficiency) should be performed. In particular, in the patient with liver disease, the practitioner should make sure that the various components of the coagulopathy are corrected prior to any invasive procedure. Coagulation factor deficiencies should be corrected with fresh frozen plasma and cryoprecipitate; significant thrombocytopenia should be treated with platelet transfusion; qualitative platelet function abnormalities should be treated with desmopresin acetate (DDAVP; 1-deamino-8-d-arginine-vasopressin). In all patients whose screening test results include abnormalities, a hematology consultation should be obtained for management of the coagulopathy in the preoperative period.

**Bleeding in Uremia.** A bleeding episode in a patient with uremia is a relatively common occasion for hematologic evaluation and deserves special consideration. Even though there is poor correlation between an abnormal bleeding time and actual bleeding in a uremic patient (21), this test remains the only available one by which to evaluate response to treatments such as intensive dialysis, desmopressin acetate, cryoprecipitate, or estrogens that may be given a trial to minimize the risk of bleeding in patients undergoing elective surgery (22).

## Polycythemia and Myeloproliferative Disorders

Hematocrit levels above the normal range (>45%) are associated with progressively diminishing blood flow to major organs (23). Moreover, the association of high hematocrit levels and thrombosis in patients with polycythemia vera (PV) is unquestionable (24). For this reason it is desirable not to perform surgery on any patient with PV whose hematocrit level is 50 or higher. If phlebotomy is necessary, it should be performed well ahead of the planned surgery in order to ensure restoration of a normal intravascular volume (plasma or plasma expanders can also be used). In patients with secondary polycythemia, the cause of the disease needs to be clarified in order to avoid risks related to hypoxia (particularly in patients with lung and congenital heart disease). Moreover, the association between high hematocrit and hyperviscosity or thrombosis is less obvious (25) and phlebotomy down to a hematocrit of about 60% should be considered for those patients with a history of hyperviscosity symptoms. The use of anticoagulants and platelet-function inhibitors in polycythemia patients needs to be weighed against the risk of bleeding because of the impairment of coagulation and platelet function often present (26).

In patients who have chronic myelogenous leukemia (CML), the main concerns are the possibility of hyperviscosity and leukostasis and the thrombocytosis that can also be present. Young patients with CML are more likely to develop leukostasis than older patients. Leukostasis occurs if the leukocrit is more than 15% and if the blood viscosity is > 4.5 centipoise (27). It is good practice to evaluate these parameters if the white blood cell count is more than 100,000/μL or if the patient has any history of headache, blurred vision, dyspnea, and so forth. Remember also that an unnecessary blood transfusion in a patient with high white blood cell count can raise the blood viscosity to a point at which leukostasis may ensue. Patients with CML may have abnormalities in platelet function similar to those in polycythemia vera.

Essential thrombocythemia is characterized by platelet counts usually higher than 1 million/μL and by the concomitant occurrence of both thrombotic and bleeding phenomena. In patients who have no history of complications related to the high platelet count, it is suggested that the platelet count be reduced to less than 1 million/μL by apheresis or medications (hydroxyurea, Anagrelide, or alpha-interferon), or both. If there is any history of thrombosis, the patient should be regarded as being at high risk of developing postoperative thrombosis and treated consequently (aminosalicylic acid is usually sufficient). If there is any history of bleeding, prophylaxis for thrombosis must be avoided. The perioperative management of patients with essential thrombocythemia is complicated by the occurrence of both hemorrhagic and thrombotic phenomena in the same patient (28).

Secondary thrombocythemias, even when the platelet count is very high, are not associated with thrombotic or hemorrhagic complications, and no specific measures are required. In order to differentiate between primary and secondary thrombocythemias, platelet function studies and bone marrow evaluation can be of help. In patients with myeloid metaplasia, a consultant should be aware of the frequent occurrence of thrombocytosis, especially after splenectomy (thrombocytosis can also occur in chronic myelogenous leukemia or essential thrombocythemia). It is advisable to reduce the platelet count to normal prior to an elective splenectomy to avoid the risks linked to excessive thrombocytosis, particularly if the patient has a history of deep vein thrombosis, coronary artery disorders, or cerebrovascular accident (29).

### Other Hematologic Malignancies

In patients with acute myelogenous leukemia or acute lymphoblastic leukemia it is best to avoid elective surgery, given the incidence of bleeding and infectious complications (30). Leukostasis presents yet another problem, in relation not only to a high white blood cell count, but also to the tendency of immature white blood cells to invade the capillary wall, with activation of the coagulation system and increased capillary permeability (31). It is important to look for evidence of leukostasis (via chest x-ray and fundoscopic and neurologic examination) in patients with acute myelogenous or lymphoblastic leukemia, regardless of the white blood cell count, and to measure viscosity and leukocrit if that cell count is > 100,000/μL. Emergency apheresis is the best way to deal with impending leukostasis if surgery has to be performed. Adequate supportive care in the form of blood products and antibiotic coverage is paramount if the surgery is to be successful (32–35).

Patients who have multiple myeloma pose not only the problems related to thrombocytopenia and leukopenia noted above but also three other potentially dangerous situations:

- The presence of hypercalcemia, which needs to be corrected before any surgical intervention; because of the surgical bleeding risk it is advisable to avoid the use of mithramycin, which can result in profound thrombocytopenia, especially in a patient with bone marrow disease (such as myeloma).
- The possibility of hyperviscosity, which is more common in patients with IgM and IgA monoclonal proteins. If the viscosity is more than 4, plasmapheresis is indicated (36).
- The possibility of platelet function abnormality or presence of coagulation factor inhibitors due to the monoclonal protein. Full study of coagulation is warranted in these patients. Moreover, whenever immunoglobulin deficiency is present, intravenous gamma globulins should be considered to minimize the risk of overwhelming infections.

In patients who have chronic lymphocytic leukemia and Hodgkin and non-Hodgkin lymphomas, immunohemolytic anemia should be ruled out. Rarely, coagulation inhibitors are also encountered. Other potential risks are the consequences of bone marrow invasion (pancytopenia) and immunosuppression, which should be dealt with according to the guidelines given above.

### Thrombocytopenia

If platelet function is normal, excessive bleeding in patients with platelet counts over 100,000/μL is rare, and counts over 50,000/μL should be adequate for most major surgery; higher levels should be present for plastic, neurologic, some ophthalmologic, and cardiac surgery. In addition to altered platelet function, fever and sepsis can also increase the possibility of bleeding, whatever the platelet count.

If a low platelet count is due to diminished production, platelet transfusion may allow a temporary rise in the platelet count sufficient to minimize the risk of bleeding and thus allow surgery to be performed. One unit of platelets should raise the count by 5–10,000; the rise must be documented by posttransfusion counts. Since the normal lifespan of newly produced platelets is 8–10 days, a platelet transfusion, consisting of platelets of various ages, cannot be expected to maintain a significant rise in the count for more than 5 days under optimal circumstances. Due to alloimmunization and splenic sequestration or consumption, the usual lifespan of transfused platelets is between several hours and 3 days.

Whenever possible and certainly prior to any elective surgery, the underlying condition causing thrombocytopenia should be treated. If the low count is due to increased destruction, the platelet transfusion is likely to be unsuccessful, and therapy must be directed at the basic disease. This is especially true in the autoimmune thrombocytopenias, in which prednisone or high-dose intravenous gamma globulins should be used prior to surgery. If the patient with thrombocytopenia is to undergo splenectomy utilizing special surgical techniques (37), the elimination of this major site of platelet destruction may permit effective use of platelet transfusion if bleeding occurs.

### Neutropenia

An absolute neutrophil count of less than 1,000/μL is associated with a higher incidence of bacterial infections. Patients with less than 500 neutrophils/μL should be hospitalized, given the high risk of overwhelming infections. In patients with severe granulocytopenia, it is best to postpone surgery or to limit it to the least invasive procedure in case of emergency (38).

### Paroxysmal Nocturnal Hemoglobinuria

All the major hematologic abnormalities that contraindicate surgery (hemolytic anemia, neutropenia, thrombocytopenia, bleeding, and thrombosis) may occur in a patient with paroxysmal nocturnal hemoglobinuria (39). If an operation is unavoidable (and often recurrent crises of abdominal pain which mimic an ileus are due to small mesenteric vein thrombosis or intraparenchymal hepatic vein thrombosis which is not amenable to surgical therapy), precautions should be taken to avoid exacerbation of hemolysis by the use of washed or frozen packed red blood cells, intravenous dextran, and androgens. The patient should be adequately hydrated to prevent renal failure. Heparin prophylaxis should be utilized to decrease the risk of thrombosis.

### Microangiopathic Hemolytic Anemias

The microangiopathic hemolytic anemias can be due to an array of causes; some are mechanical (prosthetic heart valve), postinfectious (shigellosis and different viruses causing diseases, such as hemorrhagic fever with renal syndrome), or associated with cancer or chemotherapy, and others are unknown (thrombotic thrombocytopenic purpura and some cases of hemolytic uremic syndrome). Whenever an underlying cause is suspected, that condition should be treated before any surgery is performed. When a patient has thrombotic thrombocytopenic purpura and when plasmapheresis is ineffective in controlling the signs and symptoms, splenectomy can be performed as a last attempt to modify the disease course (40). Steroids, antiplatelet drugs such as dipyridamole, intravenous dextran, and fluids should be given perioperatively in an attempt to diminish blood viscosity and platelet consumption. If possible, platelet transfusion should be avoided.

### References

1. Tudhope GR, Wilson GM. Anemia in hypothyroidism: incidence, pathogenesis and response to treatment. Q J Med 1960; 29:513–537.
2. Erslev AJ. Anemia of chronic disorders. In: Williams WJ, et al, eds. Hematology. New York: McGraw-Hill, 1983:522.
3. Mollison PL. Blood transfusion in clinical medicine. Oxford: Blackwell Scientific, 1987:471.
4. Carson JL, Spence RK, Poses RM, Bonavita G. Severity of anaemia and operative mortality and morbidity. Lancet 1988; 1:727–729.
5. Liggett SB, Schuster DP. Acute respiratory failure. In: Dunagan WC, Ridner ML, eds. Manual of medical therapeutics. Boston: Little, Brown, 1989:187.
6. Berkman SA. The spectrum of transfusion reactions. Hosp Pract [Off] 1984; 19:205–218.
7. Popovsky MA, Devine PA, Swell HF. Intraoperative autologous transfusion. Mayo Clin Proc 1985; 60:125–134.
8. Kay LA. The need for autologous blood transfusion. Br Med J 1987; 294:137–139.
9. Greenwalt TJ, Zelensky KR. Transfusion support for haemoglobinopathies. Clin Haematol 1984; 13(1):151–165.
10. Hull RD, Raskob GE, Hirsh J. Prophylaxis of venous thromboembolism: an overview. Chest 1986; 89:374S–382S.
11. Merli GJ, Martinez J. Prophylaxis for deep vein thrombosis and pulmonary embolism in the surgical patient. Med Clin North Am 1987; 71:377–397.
12. Biggs R, Macfarlane RG. Haemophilia and related conditions: 187 cases. Br J Haematol 1958; 4:1–27.
13. Rapaport SI. Preoperative hemostatic evaluation: which tests, if any? Blood 1983; 61:229–231.
14. Sack GH, Levin J, Bell WR. Trousseau's syndrome and other manifestations of chronic disseminated coagulopathy in patients with neoplasms: clinical, pathologic and therapeutic features. Medicine (Baltimore) 1977; 56:1–37.
15. Shapiro SS, Hultin M. Acquired inhibitors to the blood coagulation factors. Semin Thromb Hemost 1975; 1:336–385.
16. Suchman AL, Mushlin AI. How well does the activated partial thromboplastin time predict postoperative hemorrhage? JAMA 1986; 256:750–753.
17. Rohrer MJ, Michelotti MC, Nahrwold DL. A prospective evaluation of the efficacy of preoperative coagulation testing. Ann Surg 1988; 208:554–557.
18. Burns ER, Billett HH, Frater RWM, Sisto DA. The preoperative bleeding time as a predictor of postoperative hemorrhage after cardiopulmonary bypass. J Thorac Cardiovasc Surg 1986; 92:310–312.
19. Barber A, Green D, Galluzzo T, Chung-Hsin Ts'ao. The bleeding time as a preoperative screening test. Am J Med 1985; 78:761–764.
20. Bick RL. Disseminated intravascular coagulation and related syndromes: a clinical review. Semin Thromb Hemost 1988; 14:299–338.
21. Channing Rodgers RP, Levin J. A critical reappraisal of the bleeding time. Semin Thromb Hemost 1990; 16:1–20.
22. Fellin F, Murphy S. Hematologic problems in the preoperative patient. Med Clin North Am 1987; 71:477–487.
23. Castle WB, Jandl JH. Blood viscosity and blood volume: opposing influences upon oxigen transport in polycythemia. Semin Hematol 1966; 3:193–198.
24. Wasserman LR, Gilbert HS. Surgical bleeding in polycythemia vera. Ann NY Acad Sci 1964; 115:122–138.
25. Perloff JK, Rosove MH, Child JS, Wright GB. Adults with cyanotic congenital heart disease: hematologic management. Ann Intern Med 1988; 109:406–413.
26. Berk PB, Goldberg JD, Donovan PB, Fruchtman SM, Berlin NI, Wasserman LR. Therapeutic recommendations in polycythemia vera based on polycythemia vera study group protocols. Semin Hematol 1986; 23:132–143.

27. Lichtman MA, Rowe JM. Hyperleukocytic leukemias: rheological, clinical and therapeutic considerations. Blood 1982; 60:279–283.
28. Kutti J. The management of thrombocytosis. Eur J Haematol 1990; 44:81–88.
29. Schwartz SI. Myeloproliferative disorders. Ann Surg 1975; 182:464–471.
30. Koretz MJ, Neifeld JP. Emergency surgical treatment for patients with acute leukemia. Surg Gynecol Obstet 1985; 161:149–151.
31. Lester TJ, Johnson JW, Cuttner J. Pulmonary leukostasis as the single worst prognostic factor in patients with acute myelocytic leukemia and hyperleukocytosis. Am J Med 1985; 79:43–48.
32. Nagler A, Yoffe G, Argov S. Surgery in acute leukemia. Curr Surg 1985; 42:365–369.
33. Seligman BR, Rosner F, Ritz ND. Major surgery in patients with acute leukemia. Am J Surg 1972; 124:629–633.
34. Bjornsson S, Yates JW, Mittelman A, Holland JF. Major surgery in acute leukemia. Cancer 1974; 34:1272–1275.
35. Bishop JF, Schiffer CA, Aisner J, Matthews JP, Wiernik PH. Surgery in acute leukemia: a review of 167 operations in thrombocytopenic patients. Am J Hematol 1987; 26:147–155.
36. Ray PK, Besa E, Idiculla A, Rhodas JE, Bassett JG, Cooper DR. Efficient removal of abnormal immunoglobulin G from the plasma of a multiple myeloma patient. Cancer 1980; 45:2633–2638.
37. Watson-Williams EJ. Hematologic and hemostatic consideration before surgery. Med Clin North Am 1979; 63:1165–1189.
38. Adler AG. Medical evaluation of the surgical patient. Philadelphia: Saunders, 1985:58.
39. Braren V, Jenkins DE, Phythyon JM, Hartmann RC, Clark DA. Perioperative management of patients with paroxysmal nocturnal hemoglobinuria. Surg Gynecol Obstet 1981; 153:515–520.
40. Cuttner J. Thrombotic thrombocytopenic purpura: a ten-year experience. Blood 1980; 56:302–306.

# Preoperative Evaluation of the Patient with Cardiac Disease

HERSCHEL J. SKLAROFF, M.D.

AN EVALUATION of cardiovascular status is a prerequisite for any major surgical procedure involving general anesthesia contemplated for a patient with a heart disorder because of the strains imposed on the circulatory system and because the diseased heart is more vulnerable to this strain than the normal heart (1, 2). The major operative risk of noncardiac surgery arises primarily from coronary artery disease. Factors such as pain, apprehension, surgical stress, volume changes, increase in circulatory catecholamines, tachycardia, and hemorrhage, which are more easily tolerated by the normal heart, may have catastrophic consequences in the patient with coronary heart disease or impaired left ventricular function. The internist or cardiologist must therefore determine whether a cardiac lesion is present and, if it is, whether its severity alters the decision to operate or dictates a modification of the surgical procedure, more intensive monitoring, or type of anesthesia.

An estimate of the ability of the diseased heart to withstand the contemplated surgical procedure requires a knowledge of the risk added by the cardiac disease to the usual risk of the operation itself and of alternative therapeutic approaches. An estimate must also be made of the general outlook and life expectancy of the patient in terms of the cardiac disease. A substantial surgical risk cannot be justified if the cardiac condition precludes a survival period sufficient to warrant surgery. Furthermore, the risk of withholding surgery and its attendant dangers and disability must be carefully weighed against the risks of the proposed surgery to the cardiac patient.

In most situations, the cogency of the surgical indication is usually more important than the patient's cardiac status in determining whether an operation is to be performed. Most patients with a cardiac condition that allows for a reasonable though limited life style are capable of withstanding major surgery and general anesthesia without significantly increasing the risk. However, in the presence of serious heart disease, the evaluation of risk versus surgical indication poses a difficult problem; generous doses of experience, intuition, and divine inspiration are frequently required to arrive at the correct decision. Finally, the skill of the surgeon and the anesthesiologist must be taken into account.

## Circulatory Risks Imposed by Surgical Procedures

The most common surgical complications that adversely affect the diseased circulation are hemorrhage and shock, infection, thromboembolism, and impaired ventilation. When these complications are severe and uncontrolled they are equally fatal to patients who do and patients who do not have cardiac disorders, but when the clinical situation is precarious, the presence of heart disease might turn the tide against recovery.

Whatever the nature of the heart disease, the delivery of adequate oxygen and nutrients via the coronary arteries to the myocardium is crucial. Hemorrhage, anemia, hypotension, and hypoxemia, which might be tolerated by a normal heart, may result in serious and irreparable damage to and fatal disturbance of the diseased heart.

Infection, by imposing circulatory and metabolic demands incapable of being met by the im-

---

From the Division of Cardiology, Department of Medicine, Mount Sinai School of Medicine, New York, NY 10029. Address reprint requests to the author at 1175 Park Avenue, New York, NY 10128.

paired heart, continues to be a leading cause of heart failure in patients with cardiac disorders. Despite great advances in diagnostic techniques, including radionuclide scanning and the introduction of a vast array of antimicrobials, infection continues to exact a substantial morbidity in surgical patients. The most common perioperative infectious process other than wound infection is pneumonia, caused by hypoventilation from mucous plugs, atelectasis, narcotic-induced respiratory depression and cough suppression, painful inspiration, and poor pulmonary drainage from prolonged bed rest and from failure to frequently turn the bedridden patient.

Thromboembolic disease continues to plague the surgical patient, especially in the presence of obesity, prolonged bed rest, congestive heart failure, hip fracture, lengthy surgery, cancer, and other hypercoagulable states. The sudden onset of hemoptysis, pleuritic chest pain, dyspnea, tachypnea, tachycardia, arrhythmia, cyanosis, hypotension, or heart failure should alert the physician to the possibility of pulmonary thromboembolism and to the institution of immediate diagnostic and subsequent therapeutic intervention. The use of sophisticated diagnostic techniques, such as pulmonary perfusion scanning, pulmonary arteriography, venography, plethysmography, and duplex sonography, coupled with the use of anticoagulants and thrombolytics, such as heparin, Coumadin, streptokinase, and tissue plasminogen activator, give promise of greatly reducing morbidity from this dreaded complication.

Disturbances in ventilation, such as atelectasis, mucous plugging, pneumonia, pneumothorax, pleural effusions, emphysema, and pulmonary embolism, are more frequent and more serious in the cardiac than in the noncardiac patient, due perhaps to chronic pulmonary congestion and "stiff lungs" from left-sided heart failure or to the frequency of emphysema in the elderly. Despite the greater hazard of these surgical complications in the cardiac patient, their incidence is relatively low, and the severity of the complication itself, rather than the presence or absence of cardiac disease, usually determines the outcome, except in borderline cases.

## Circulatory Risks Imposed by Anesthesia

The risk of anesthesia is related to the dangers of hypoxemia, vagal reflexes, hypercarbia, sharp changes in blood pressure, and the induction of arrhythmias. Hypoxemia may be related to low concentrations of oxygen in the anesthetic mix or to mechanical causes interfering with the respiratory airway. The damaged heart, especially the heart with diminished coronary reserve, is particularly susceptible to hypoxemia, which may induce myocardial necrosis or predispose to ventricular fibrillation and cardiac arrest. The induction of anesthesia, especially in the patient with cardiac disease, may result in benign or malignant arrhythmias. Hypotension, which can severely reduce coronary blood flow, may accompany spinal anesthesia, large doses of barbiturates, and inhalational anesthetics which induce marked peripheral vasodilatation. The excitement and struggling associated with the induction of anesthesia may increase the work of the heart and be especially detrimental to patients with coronary artery disease.

**Cardiovascular Effects of Anesthesia.** The choice of anesthetic agent and technique is largely that of the anesthesiologist, but the internist and cardiologist must have a basic understanding of the cardiovascular effects of anesthesia to render optimum patient care. The complication rate of anesthesia in the cardiac patient does not usually depend on the choice of regional versus general anesthesia, or on the agent used, but rather on the anesthesiologist's skill, familiarity with the technique and anesthetic used, and, above all, careful monitoring of the patient (3–9).

Most general anesthetics depress myocardial contractility to a greater or lesser degree, but often stimulate the sympathetic nervous system, thereby offsetting that suppression. Thus, halothane may induce hypotension, whereas cyclopropane and ether, for example, are less likely to do so. Light halothane anesthesia provokes minimal cardiac depression and vasodilatation. Enflurane anesthesia is also well tolerated by the cardiac patient and is minimally arrhythmogenic. Some of the newer anesthetic combinations, such as droperidol and fentanyl, or nitrous oxide and pancuronium bromide, produce little myocardial suppression. However, a disadvantage of balanced anesthesia is the inability to lighten the level of anesthesia quickly, as can be done when inhalational anesthesia is the sole technique.

An excellent option for anesthesia in poor-risk cardiac patients is the use of nitrous oxide, a narcotic, and a muscle relaxant such as succinylcholine, curare, or pancuronium. Fentanyl has replaced morphine because it produces less hypotension. With this combination, myocardial suppression is rare, hypotension unusual, respiration easily controlled with intubation, and narcotic effects rapidly reversed with narcotic antagonists.

Intravenous thiopental or benzodiazopines are commonly used for induction of anesthesia. Barbiturates may cause myocardial suppression and hypotension and are therefore poorly tolerated by the cardiac patient. Ketamine may be substituted, because it exerts a mild sympathomimetic effect. It may also be used in patients with cyanotic congenital heart disease, because its hypertensive effect may diminish right-to-left shunting. Regardless of the agent employed in general anesthesia, a 50% oxygen concentration in the inspired mixture is desirable in the cardiac patient.

An alternative to general anesthesia is the use of spinal epidural or regional anesthesia, as these modes are free of myocardial depression. However, the hypotension induced by sympathetic blockade may be especially deleterious to the cardiac or hypovolemic patient, and careful blood pressure control is important to the outcome. Nerve block anesthesia is satisfactory for operations on the extremities, whereas epidural or spinal anesthesia is ideal for surgery of the prostate gland, spine, pelvis, and lower extremities. Local anesthetics without epinephrine may be used for minor surgery but can be absorbed systemically and produce negative inotropic and chronotropic cardiac effects. There is probably less postoperative heart failure with spinal and epidural anesthesia than with general anesthesia.

**Techniques of Anesthesia.** Induction of anesthesia may cause temporary hypotension, hypoxemia, hypercarbia, and arrhythmias. Rapid intubation in the well-anesthetized patient is therefore mandatory. Endotracheal aspiration, if necessary, should be done only briefly, because similar undesirable side effects may occur. Extubation should be done only when full respiratory function is regained; extra caution is required in patients with pulmonary and renal disease, myopathies, and biochemical aberrations.

In patients with severe underlying heart disease, cardiac function must be carefully monitored (cardiac rate and rhythm, direct recording of arterial pressure, blood gas determination, and measurement of pulmonary wedge pressure, venous pressure, and cardiac output with the aid of the Swan-Ganz catheter and the thermodilution method). In my experience, however, the use of the Swan-Ganz catheter, which is not without risk, is excessive, and the catheter is often inserted when only the vaguest history of heart disease is elicited or when minor cardiac abnormalities are detected. Furthermore, it frequently fails to provide useful data because of a variety of technical problems.

**Arrhythmias during Anesthesia.** Alterations in cardiac rhythm are extremely common during anesthesia and are likely to occur in patients with coronary artery disease and calcific aortic stenosis (10, 11). Most arrhythmias require no therapy and respond to minor adjustments of anesthesia management, such as correction of hypoxemia, hypercarbia, and the depth of anesthesia. When arrhythmias are not corrected by these means, they should be treated with the appropriate antiarrhythmic agent if judged potentially serious.

## Management and Preparation of the Patient

Careful preparation of the cardiac patient for surgery may diminish the frequency and seriousness of intraoperative and perioperative complications (12). Since hypoxemia and hypotension cause most of these complications, every effort must be made to prevent their occurrence.

A complete history should be taken, with special attention to recent medication, response to drugs and drug sensitivity, and the use of corticosteroids and antihypertensive therapy. A history of paroxysmal or chronic arrhythmias, angina pectoris, recent or past myocardial infarction, or manifestations of heart failure is important with respect to management and prognosis. A complete physical examination should be performed, as well as examination of the body area involved in the contemplated surgical procedure. An electrocardiogram, chest roentgenogram, complete blood count, prothrombin time, partial thromboplastin time, and complete battery of routine chemistry studies should be obtained to assess correctable abnormalities and serve as a baseline to compare with possible future alteration.

During the preoperative period, congestive heart failure should be treated without inducing hypovolemia, electrolyte imbalance, or digitalis intoxication. Anemia should be corrected to a hemoglobin of at least 100 g/L, polycythemia corrected to diminish bleeding and thrombosis, arrhythmias controlled, pacemakers inserted, antihypertensives adjusted to achieve good control, electrolyte disturbances corrected, anticoagulant therapy modified or discontinued, and pulmonary function made optimal. The choice of anesthesia and appropriate monitoring modes should be discussed with the surgeon and anesthesiologist.

If impaired pulmonary function is present, patients should obviously desist from smoking.

Antibiotics, postural drainage, bronchodilators, and corticosteroids may diminish postoperative pulmonary complications. If these patients do not exhibit dyspnea, tachypnea, cyanosis, or significant cough at rest, there is a lesser likelihood of intraoperative or postoperative pulmonary complications, and routine pulmonary function tests may not be necessary.

Patients with bronchial asthma present a more difficult problem. They should be rendered wheeze-free, when possible, with the use of bronchodilators such as theophylline, Proventil, and Atrovent and by inhalational and, if necessary, systemic corticosteroids. However, bronchodilators may alter the patient's vital signs in the perioperative period, thus confusing the clinical picture. I therefore discontinue all such medication 24 hours prior to surgery and begin systemic corticosteroids two days prior to surgery, usually Decadron in a dose of 6–8 mg daily. This approach has met with uniform success.

Prophylaxis against and treatment of heart failure should begin with the avoidance of sodium-containing fluids, including blood, plasma, and amino acid infusions. Should such fluids be essential, special caution as to rate and quantity of administration should be urged. Blood transfusion should always be given concomitantly with intravenous furosemide. Occasionally, preoperative monitoring of venous and pulmonary wedge pressures may be helpful in recognizing heart failure, but such monitoring is no substitute for auscultation of the heart and lungs. Some patients with heart disease may be intolerant of certain positions required intraoperatively which result in changes of pulse and blood pressure. Preoperative observation of vital signs in various positions should be performed.

If heart failure is present preoperatively, treatment with digitalis, diuretics, and oxygen is necessary. Rapid atrial fibrillation should be controlled with digoxin, and thyroid function tests should be obtained. Surgery should be delayed as long as required to obtain optimum cardiac performance but not so long as to jeopardize the desired outcome.

**Other Preoperative Measures.** There are many regimens for the perioperative administration of insulin, such as giving one-third to one-half the usual dose preoperatively. I prefer, however, to give only regular insulin on a sliding scale by frequent monitoring of the easily, rapidly obtainable blood sugar level and continue to do so until the patient returns to the preoperative state and is eating his or her usual diet. This method avoids the potentially large swings in blood sugar that are associated with stress, fever, infection, and intravenous fluids.

Corticosteroids, usually in the form of Solucortef, should be given in stress doses if they were previously taken to avoid relative or absolute adrenal insufficiency.

Antihypertensive medication may be continued if necessary, especially if withdrawal would be associated with rebound hypertension and tachycardia, as might be the case with clonidine and the beta blockers. Dosage may change with clinical circumstance, and careful blood pressure monitoring of the patient in both the supine and upright positions is helpful in avoiding overtreatment.

Anticoagulants may be discontinued preoperatively with little risk and restarted when hemorrhage is deemed unlikely. The patient with a mitral valve prosthesis should discontinue Coumadin 4–5 days prior to surgery and heparin should be given up to 12 hr preoperatively. Postoperatively, heparin and coumadin are given when it is safe, the heparin being discontinued when the prothrombin time is within the therapeutic range. Correction of the prothrombin time with vitamin K should be avoided, if possible, because subsequent Coumadin resistance is frequently encountered. Aspirin should be discontinued 10–14 days prior to surgery and the bleeding time obtained. Clotting abnormalities may be corrected with the appropriate therapy: protamine, vitamin K, desmopressin acetate, fresh frozen plasma, platelets, or fibrinogen.

Antibiotic prophylaxis for congenital heart disease, rheumatic valves, prosthetic valves, and mitral valve prolapse, and for patients with a previous history of infectious endocarditis is recommended before and after surgery. Amoxacillin for operations above the diaphragm and amoxacillin with gentamicin for operations below the diaphragm is standard therapy. Vancomycin may be substituted when penicillin allergy exists. Antibiotics should be continued until the risk of surgical infection abates.

## The Operation

Minor procedures such as dental extractions, incision and drainage of abscesses, suturing, and herniorrhaphy can be performed as outpatient procedures and pose no risk for the cardiac patient. Early surgical intervention for operable malignant neoplasms is obviously indicated when there is a chance of cure and the cardiac condition is not in the terminal stage.

When a surgical emergency arises, if the condition requiring surgery is otherwise fatal, an operation is essential even if the patient has heart disease. Such conditions include

- acute and massive hemorrhage
- infarction or perforation of a viscus
- most dissecting and all ruptured aneurysms
- strangulated hernia
- acute appendicitis
- torsion of the ovary or testicle
- certain types of acute intestinal obstruction
- acute cholecystitis or acute diverticulitis, especially in the patient with diabetes or vasculitis

In patients who have acute cholecystitis or acute diverticulitis, antibiotic therapy may obviate the need for emergency surgery of the condition if clinical observation discloses regression of clinical features. Gangrene of an extremity resulting from thrombosis, embolism, or infection must be surgically treated if possible.

Elective surgery in patients with cardiovascular disease may pose problems that tax the clinical judgment of the medical consultant, especially in patients who have recently had myocardial infarction (3–7). The operative risk in patients with underlying heart disease depends on the type of surgery. Herniorrhaphy and transurethral prostatectomy pose little risk, whereas cardiac patients undergoing cholecystectomy, subtotal gastrectomy, or bowel resection have perhaps twice the mortality of patients without heart disease. In patients with previous myocardial infarction, major vascular surgery, thoracic surgery, and upper abdominal surgery have a two- to three-fold risk of death (4, 7). Emergency surgery in the cardiac patient entails greatly increased mortality compared to elective surgery. Perioperative cardiac mortality may increase with the duration of the surgical procedure (4).

Urinary obstruction due to prostatic hypertrophy in patients with coronary artery disease is one of the most common conditions for which an operation is performed. These patients are usually elderly. Coronary occlusion in the perioperative period is relatively common. Prostatectomy is indicated if the patient has

- episodes of complete urinary obstruction
- severe retention with secondary infection or impairment of renal function
- symptoms that are intolerable

Of note is that heart failure may be markedly improved and the frequency of angina pectoris diminished when the obstruction is relieved. If surgery is deemed exceedingly risky or when life expectancy is short, catheter drainage is an acceptable alternative to prostatic surgery.

Operations on the biliary tract are frequently necessary in the cardiac patient. Acute suppurative or gangrenous cholecystitis in men and in patients with either diabetes or vasculitis usually requires relatively urgent surgery. Obstructive jaundice almost always demands operative intervention, but recent advances, including ERCP with papillotomy, lithotripsy, chemical dissolution of gallstones, and choledochotranshepatic stent placement, have achieved the desired result without subjecting such patients to the risks of surgery under general anesthesia. A history of frequent attacks of biliary colic not amenable to these techniques requires surgery, even in the high-risk cardiac patient. In the well-compensated patient with heart disease, early surgical intervention may be preferable to waiting—the patient will be older and presumably sicker. Cholecystectomy is generally well tolerated and of little risk. The mortality of cholecystectomy via peritoneoscopy versus conventional laparotomy remains to be seen, but the use of peritoneoscopy holds promise.

Other operations, such as hemorrhoidectomy, hysterectomy, and oophorectomy, pose little risk.

The patient with a fractured hip should be operated on as soon as possible, because for elderly patients especially, prolonged bed rest is tantamount to a fatal outcome. In these patients, the risk of pulmonary embolism is formidable, and anticoagulation therapy, barring contraindication, is mandatory. Spinal or regional anesthesia is usually preferable, and the earlier the patient walks, the better.

## Assessment and Reduction of Cardiac Risk

Several cardiac risk factors have been shown to be associated with perioperative cardiac complications (3):

- age over 70 years
- myocardial infarction within 6 months of surgery
- previous pulmonary edema
- class III or IV or unstable angina pectoris
- previous myocardial infarction with a subsequent ventricular ejection fraction of less than 35%
- presence of a third heart sound or jugular venous distention
- significant aortic stenosis determined by physical examination and Doppler echocardiography

- arrhythmias other than premature atrial contractions
- more than five premature ventricular beats per minute in patients with previous myocardial infarction or left ventricular dysfunction from any cause
- intrathoracic, intraperitoneal, or aortic surgery
- emergency operation
- poor general medical condition

Correction of these factors where posible would presumably diminish surgical risk.

Factors found not to be important in determining surgical outcome were controlled diabetes mellitus, the presence of a fourth heart sound, controlled hypertension, hyperlipidemia, peripheral vascular disease, and chronic, stable class I or II angina pectoris.

**Coronary Heart Disease.** The surgical risk is somewhat greater in the patient with coronary artery disease and therefore the surgical indication should be unequivocal. Surgical mortality and postoperative myocardial infarction are high when surgery is performed within 6 months of a previous myocardial infarction (5, 6, 13). However, despite the increased surgical risk, most patients with clinical coronary artery disease tolerate major surgery quite well, provided adequate measures are taken to prevent myocardial ischemia. Avoidance of preoperative excitement, hypoxemia, hypotension, and deep anesthesia is mandatory. An electrocardiogram should always be taken preoperatively to exclude recent myocardial infarction, and postoperatively to exclude an intraoperative ischemic event (7, 14). Emergency lifesaving surgery in the coronary heart patient should obviously be performed, using the best hemodynamic monitoring systems available. When surgery is semielective, as in the patient with a potentially resectable tumor, a delay of 4–6 weeks is acceptable following myocardial infarction if the patient is at relatively low risk (15, 16). However, when the patient is at high risk from class III or IV or unstable angina pectoris or has had a previous myocardial infarction and currently has angina pectoris, I believe that coronary arteriography should be performed and coronary bypass or percutaneous coronary angioplasty should be done if indicated and feasible prior to surgery (17, 18).

Patients who are maintained on an antianginal regimen should have such therapy continued in the perioperative period. Beta blockers should be given preoperatively and postoperatively. Inderal may be administered intravenously in doses of 1–2 mg every 4–6 hr or as a constant infusion at a rate of 0.01–0.5 mg/min after a loading dose of 5–10 mg over 60 min. This regimen may be continued until the patient is capable of oral intake. When nitrate therapy is required, a nitro patch or intravenous nitroglycerine may be used. Occasionally the use of sublingual nifedipine or intravenous verapamil may be necessary for angina, especially when due to coronary vasospasm.

**Valvular Heart Disease.** Patients with valvular heart disease require perioperative support when undergoing noncardiac surgery; a knowledge of the cardiac risk associated with the valvular defect, the hemodynamic significance of the lesion, assessment of left ventricular function, the effect of alteration of heart rate and rhythm, and the interaction of medications used in the perioperative state must all be taken into account. A careful clinical history, with special reference to dyspnea, cardiac asthma, orthopnea, exercise tolerance, and previous episodes of heart failure, is essential for evaluation of cardiac function.

The operative mortality rises dramatically from each lower New York Heart Association classification to each higher classification, approaching 70% in class IV patients. It follows that successful preoperative treatment of heart failure may significantly reduce the operative risk. Attention must be paid to electrolyte and volume changes and avoidance of digitalis intoxication. As with coronary artery disease, emergency surgery demands hemodynamic monitoring and doing whatever is indicated to enhance cardiac performance within the constraints of time and the type of surgery. If the contemplated operation is semielective, corrective surgery—for example valve replacement or balloon valvuloplasty—may be performed as warranted.

*Regurgitant and Stenotic Lesions.* In general, regurgitant valvular lesions are better tolerated than stenotic lesions.

Significant aortic stenosis has been associated with a high operative mortality rate. The stroke volume is often fixed, and the patient may not be able to significantly increase his or her cardiac output. Hypovolemia and hypotension are poorly tolerated, and left ventricular compliance is also diminished owing to hypertrophy. These features mandate careful monitoring of volume and pulmonary capillary wedge pressure. Atrial fibrillation is especially hazardous because of the loss of "atrial kick" and diminished left ventricular filling time. Reduction of rapid ventricular response and restoration to sinus rhythm should be attempted. When possible, severe symptomatic aortic stenosis should be corrected with valve replacement, or balloon valvuloplasty when replacement is not feasible.

Chronic aortic regurgitation subjects the left ventricle to an additional volume load with eventual impairment of left ventricular function. Factors such as hypertension that increase afterload should be prevented, and those that cause bradycardia, thereby increasing aortic regurgitation, should be avoided. Volume changes in these patients are not as deleterious as in patients with aortic stenosis, and except in severe cases, invasive hemodynamic measurement is usually not required (19).

In patients with mitral stenosis, control of heart rate and intravascular volume is paramount. Elevated left atrial pressure is required to maintain left ventricular filling and cardiac output; hypovolemia must therefore be avoided. Fluid overload is especially hazardous because it may result in pulmonary edema. Tachycardia, especially from rapid atrial fibrillation, diminishes diastolic filling time and thus cardiac output; it can be controlled with digoxin and the addition of Inderal if necessary. The value of rate control with Inderal far outweighs its negative inotropic potential in patients with mitral stenosis and good left ventricular function. Small changes in blood volume may result in marked hemodynamic alterations, thus monitoring of pulmonary wedge pressure is mandatory in patients with significant mitral stenosis.

Mitral regurgitation poses few perioperative problems unless left ventricular dysfunction is severe, as manifest by signs and symptoms of heart failure. The left atrium acts as a sump for the regurgitant blood flow and so protects the pulmonary vasculature. Vasodilatation may improve the hemodynamics of mitral regurgitation by reducing left ventricular afterload, thereby increasing cardiac output. Significant increases in blood pressure and blood volume should be avoided because they will increase the degree of mitral regurgitation.

*Hypertrophic obstructive myocardiopathy* is characterized by a dynamic subaortic valvular pressure gradient and by impaired left ventricular filling. The lesion is frequently unsuspected or misdiagnosed, and echocardiography should be performed when cardiac murmurs or cardiac symptoms of obscure origin are detected, the echocardiogram being virtually diagnostic. Drugs that increase myocardial contractility, such as digitalis and beta-adrenergic catecholamines, should be avoided; they increase the left ventricular pressure gradient. Any situation that diminishes left ventricular preload, such as hypovolemia, hypotension, and tachycardia, will augment the subaortic valvular pressure gradient, thereby reducing cardiac output. Corrective measures include fluid replacement and the use of peripheral vasoconstrictors; supraventricular tachycardias are best treated with beta or calcium channel blockade. Digoxin, although relatively contraindicated, may be given to treat rapid atrial fibrillation that occurs in patients who have idiopathic hypertrophic subaortic stenosis because the benefit of rate reduction may offset the risk of increasing the pressure gradient. In these circumstances, however, careful monitoring of the cardiac output and hemodynamic state is essential.

*Valve Prostheses and Anticoagulants.* A major problem for patients with prosthetic heart valves who receive anticoagulants to prevent thromboembolism is the necessary discontinuation of anticoagulation therapy in the perioperative period. Anticoagulants may be safely discontinued three days pre- and postoperatively without complication, except in patients with a caged valve prosthesis in the mitral position, in whom a slight increase in thromboembolic events has been observed (20).

Patients with aortic valve prostheses may undergo surgery when the prothrombin time is within 2 sec of the control value. When postsurgical hemostasis is secure, heparin and Coumadin in full therapeutic dosage are begun, heparin being discontinued when the prothrombin time is in therapeutic range.

Patients with mitral valve prosthesis should follow the same protocol, except that hospitalization for heparin therapy is required during Coumadin withdrawal, the heparin being discontinued 12 h prior to surgery.

With an acknowledged paucity of data, I have been treating these patients with Persantine in the perioperative period in an attempt to further reduce thromboembolic events. Bleeding is not increased, as it is with aspirin.

*Avoiding Infectious Endocarditis.* Finally, antibiotic prophylaxis against infectious endocarditis is required for patients with rheumatic or other acquired valvular heart disease, prosthetic valves, most congenital cardiac malformations, hypertrophic obstructive myocardiopathy, mitral valve prolapse with mitral regurgitation, and previous history of bacterial endocarditis.

**Patients with Heart Failure.** Heart failure is a major determinant of operative risk. The incidence of postoperative pulmonary edema and mortality increases directly with a rise in the patient's New York Heart Association classification (21, 22). The surgical outcome appears to be related to the immediate preoperative state; therefore, treatment of heart failure should be vigorous and should be undertaken well in advance of the contemplated surgery. Note, however, that more than 50% of the cases of perioperative pulmonary edema occur in patients without an antecedent cardiac history; this edema seems to be related to advanced age, abnormal preoperative electrocardiogram, and intra-abdominal and intrathoracic surgery (21). Those patients who develop postoperative pulmonary edema do so within the first few hours of anesthesia (23), presumably owing to volume overload, anesthetic-induced myocardial suppression, cessation of positive pressure ventilation allowing increased pulmonary blood flow, and postoperative hypertension causing increased left ventricular afterload and heart failure. Maintenance therapy of heart failure must be continued in the postoperative period and modified as necessary. Intravenous medication should be administered when circumstances preclude the peroral route.

**Hypertension.** Operative procedures are well tolerated by patients whose hypertension is mild to moderate (less than 110 mmHg diastolic) and uncomplicated by congestive heart failure, renal failure, coronary heart disease, or previous cerebral vascular accident. When hypertension is severe, control is necessary, and if time allows, a normotensive state should be achieved. Overtreatment may be avoided by careful monitoring of orthostatic blood pressure.

Prolonged bed rest is an excellent antihypertensive, and I have seen many patients whose lengthy preoperative hospitalization allowed a trial of discontinuation of drug therapy in the belief that a drug-free perioperative state is preferable where possible. Many of these patients required no medication while hospitalized but usually required resumption of therapy after discharge.

I have also seen many patients whose antihypertensive regimen was continued in the postoperative period, resulting in severe hypotension when the patient was allowed out of bed. I therefore order half the usual dose of medication in the early postoperative period and carefully monitor supine and orthostatic blood pressures, modifying the therapy as needed.

Difficulty is often encountered in the use of antihypertensive medication in the perioperative period, especially when medication cannot be taken orally. When hypertension is well controlled, therapy should be given up to and including the day of surgery to help ensure intraoperative hemodynamic stability. Postoperatively, intravenous drug therapy may have to be given when the oral route is precluded.

Diuretic therapy may cause hypovolemia and hypokalemia, which should be corrected. Methyldopa and Inderal may be given intravenously to prevent withdrawal phenomena. Discontinuation of clonidine may result in a syndrome simulating pheochromocytoma when the preoperative dose exceeds 1.2 mg daily. The syndrome can be prevented by the use of a transdermal clonidine patch. Intravenous labetalol, Aldomet, and nitroprusside are also effective and can be used as well in patients whose hypertension was previously controlled with the angiotensin-converting enzyme inhibitors. Finally, calcium channel blocking agents are now frequently used as antihypertensives, and when necessary, intravenous verapamil can be administered by constant infusion, as can many of the medications already mentioned.

**Peripheral Vascular Disease.** Myocardial infarction is a significant cause of morbidity and mortality following surgery for peripheral vascular disease and abdominal aortic aneurysm. Approximately 30% of such patients have significant coronary artery disease perhaps warranting preoperative revascularization or percutaneous angioplasty when the operation is elective (24).

The presence of ischemic heart disease may not be manifest by angina pectoris because of these patients' limitation of normal activity. Approaches to identifying patients with significant coronary artery disease include treadmill electrocardiographic stress testing and exercise radionuclide angiography (25–27), dipyridamole thallium scintography (28–30), and ambulatory electrocardiographic monitoring to detect silent ischemia (31). If any of these tests are positive, identifying the patient as at high risk, coronary arteriography should be performed, followed by coronary artery revascularization for significant lesions. The risk of a coronary event is negligible in the absence of myocardial ischemia.

Patients with asymptomatic carotid bruits are not at increased risk of sustaining perioperative stroke from elective noncardiac surgery, with the possible exception of the patient with a critical cerebral vascular lesion who suffers severe hypotension and shock.

**Conductive System Disturbances and Pacemakers.** A cardiac pacemaker should be inserted preoperatively when the patient meets the criteria for insertion of a permanent pacemaker. If time does not allow, a temporary pacemaker should be placed, followed by implantation of a permanent pacemaker postoperatively. Although there is some controversy regarding the indications for a permanent pacemaker, the following conditions are generally accepted:

- complete heart block
- Mobitz type II heart block
- symptomatic sinus bradycardia
- sick sinus syndrome with documented symptomatic bradycardia
- bifascicular block with unexplained syncope

Asymptomatic chronic bifascicular block requires no pacemaker, because the possibility of intraoperative complete heart block is remote. A means for temporary cardiac pacing must nevertheless be always available.

Patients with permanent pacemakers should have them evaluated for normal pacing and sensing function prior to surgery.

The most significant potential problem associated with surgery in patients with pacemakers is the inhibition of demand pacemakers by electrocautery-induced electromagnetic energy (32). This problem may be prevented by placing the ground plate as far away from the pacemaker pulse generator as possible, avoiding use of cautery near the pacemaker, programming the pacemaker to a fixed rate mode, and limiting the frequency and duration of electrocautery to 1.0-sec bursts 10 sec apart (33).

Electrocardiographic monitoring is usually unobtainable during periods of cautery-induced electrical interference, and alternate methods of heart rate monitoring should be employed.

Because of the great number of pacemakers currently in use, the type of pacemaker, model number, response to electrical interference, and programming method should be known prior to surgery.

## References

1. Friedberg CK. Surgical procedures in the cardiac patient, chap. 4–9. In: Diseases of the heart, 3rd ed. Philadelphia: Saunders, 1966.
2. Braundwald E. General anesthesia and noncardiac surgery in patients with heart disease, chap. 54. In: Heart disease. Philadelphia: Saunders, 1980.
3. Goldman L, Caldera DL, Southwick FJ, et al. Cardiac factors and complications in noncardiac surgery. Medicine 1978; 57:357.
4. Steen PA, Tinker JH, Tarhan S. Myocardial reinfarction after anesthesia and surgery. JAMA 1978; 239:2566.
5. Arkins R, Smessaert AA, Hicks RG. Mortality and morbidity in surgical patients with coronary heart disease. JAMA 1964; 190:485.
6. Topkins MS, Artusio JF Jr. Myocardial infarction and surgery: a five year study. Anesth Analg 1964; 43:716.
7. Tarhan S, Moffitt EA, Taylor WF, Guliania ER. Myocardial infarction after general anesthesia. JAMA 1972; 220:1491.
8. Sapala J, Ponka J, Duvernoy W. Operative and nonoperative risks in the cardiac patient. J Geriatr Soc 1975; 23:529.
9. Sapala J, Arkins R, Tinker J. Perioperative myocardial infarction. Semin Anesth 1982; 1:253.
10. Katz RL, Bigger JT Jr. Cardiac arrhythmias during anesthesia and operation. Anesthesiology 1970; 33:193.
11. Vanik PE, Davis HS. Cardiac arrhythmias during halothane anesthesia. Anesth Analg 1968; 47:299.
12. Abrams LM, Chambers PA. Preoperative management. In: Kaplan JA, ed. Cardiac anesthesia. New York: Grune & Stratton, 1979:169.
13. Mattingly TW. Patients with coronary artery disease as a surgical risk. Am J Cardiol 1967; 13:274.
14. Driscoll AC, Hokika JH, Elsten BE, Proger S. Clinically unrecognized myocardial infarction following surgery. N Engl J Med 1961; 264:633.
15. DeBusk R, Blomquist C, Kouchouckos M, et al. Identification and treatment of low risk patients after acute myocardial infarction and coronary artery bypass graft surgery. N Engl J Med 1986; 314:161.
16. Goldman L, Caldera PL, Mussbaum SR, et al. Multifactorial index of cardiac risk in noncardiac surgical procedures. N Engl J Med 1977; 247:849.
17. McCollum CH, Garcia-Rinaldi R, Graham JM, DeBakey ME. Myocardial revascularization prior to subsequent artery disease. Surgery 1977; 81:302.
18. Mahar LJ, Steen PA, Tinker HH, Vlietstra RE, Smith HC, Pluth JR. Perioperative myocardial infarction in outpatients with coronary artery disease with and without aorta–coronary artery bypass grafts. J Thorac Cardiovasc Surg 1978; 76:533.
19. Hirschfeld J Jr. Surgery in the patient with valvular heart disease. In: Goldman P, ed. Medical care of the surgical patient. Philadelphia: Lippincott, 1982:99.
20. Tinker JH, Turhan J. Discontinuing anticoagulant therapy in surgical patients with cardiac valve prostheses. JAMA 1978; 239:738,
21. Goldman L, Caldera D, Southwick F, et al. Cardiac risk factors and complications in noncardiac surgery. Medicine 1978; 57:257.
22. Skinner JF, Pearce ML. Surgical risk in the cardiac patient. J Chronic Dis 1964; 17:57.
23. Cooperman LH, Price HL. Pulmonary edema in the operative and postoperative period: a review of 40 cases. Ann Surg 1970; 172:883.
24. Hertzer M, Beven E, Young J, et al. Coronary artery disease in peripheral vascular patients. Ann Surg 1984; 199:223.
25. Ardus E, Baum P, Cutler B. The ischemic exercise test in patients with peripheral vascular disease. Arch Surg 1984; 119:780.
26. Cutler B, Wheeler HB, Paraskas J, Cardalla P. Assessment of operative risk with electrocardiographic exercise testing in patients with peripheral vascular disease. Am J Surg 1979; 137:484.
27. Sain K, Patel K, Doctor U, Peck S. Preoperative cardiac screening before peripheral vascular operations. Am Surg 1985; 51:77.
28. Boucher CA, Brewster PC, Parling RC, Okada RD, Strauss HW, Pohost GM. Determinations of cardiac risk by dipyridamole-thallium imaging before peripheral vascular surgery. N Engl J Med 1989; 312:389–394.

29. Leppo J, Plaja J, Gionet M, Tumolo J, Paraskos JA, Cutler BS. Noninvasive evaluation of cardiac risk before elective vascular surgery. J Am Coll Cardiol 1987; 9:269–276.

30. Eagle KE, Singer PE, Brewster PC, Darling RC, Mulley AG, Boucher CA. Dipyridamole-thallium scanning in patients undergoing vascular sugery: optimizing preoperative evaluation of cardiac risk. JAMA 1987; 257:2185–2189.

31. Cohn PF. Silent myocardial ischemia: prevalence and prognosis. Am J Med 1985; 79(Suppl 3A):2–6.

32. Goldman L, Wolf M. The heart and circulation. In: Vandam L, ed. To make the patient ready for anesthesia: medical care of the surgical patient, 2nd ed. Menlo Park, CA: Addison-Wesley, 1984:1–20.

33. Simon A. Perioperative management of the pacemaker patient. Anesthesiology 1977; 46:127.

# Preoperative Evaluation of the Patient with Endocrine Disease

Paul L. Gilbert, M.D.

Diabetics, patients with thyroid disease, and those ingesting steroids are sufficiently common to present frequent problems in preoperative evaluation. Barely 2% of obese patients are overweight because of endocrine malfunction. However, obesity is common and associated with multiple medical problems. Therefore, a brief discussion of the subject is warranted.

### Obesity

Moderate to severe obesity may be associated with hypothalamic lesions, insulinoma and hyperinsulinism, hypogonadism, and Cushing's disease. Hypopituitarism and hypothyroidism may be associated with modest weight gains but are seldom associated with significant obesity. The clinical presentations of these morbid states permits proper diagnosis and therapy.

Obesity is associated with an increased incidence of several medical conditions of which the consulting internist must be aware:

- Gall bladder disease, which is 3 to 4 times more common than in the nonobese, perhaps related to an altered ratio of bile salts to lecithin and cholesterol in bile.
- Esophageal reflux appears to have an increased incidence.
- Hypertension, non-insulin-dependent (type 2) diabetes mellitus, hypertriglyceridemia, and hypercholesterolemia (increased LDL and decreased HDL fractions) occur in higher coincidence than in the nonobese. Obesity is not an independent variable in the genesis of atherosclerotic heart disease. Obesity is associated with increased insulin levels, increased insulin response to secretagogues, decreased tissue sensitivity, and receptors to and for insulin. These factors are manifest by an increase in daily insulin production in the obese. Hyperinsulinemia increases appetite, which further contributes to weight gain.
- Functional residual capacity decreases in mild to moderate obesity; other lung volumes are relatively well preserved. As weight increases further, the tissue mass of the chest and abdomen and altered diaphragmatic function increase the work of breathing, encroachment on other lung volumes occurs, a restrictive ventilatory defect develops, and the existing bibasilar microatelectasis increases, resulting in greater ventilation-perfusion mismatches. All these factors contribute to the hypoxia that occurs in many extremely obese patients. Although these patients have normal or increased respiratory responses to hypoxia, their responses to hypercarbia are blunted. The incidence of sleep apnea and alveolar hypoventilation syndromes is some 25 times more common in severely obese patients. (Sleep apnea may occur in as many as 30% of such patients. A history of excessive daytime drowsiness, loud snoring, insomnia, or restless sleep in such a patient should raise the question whether this diagnosis applies.)
- Altered uptake and distribution of drugs; obese patients require larger doses of fat-soluble drugs (which turn over less rapidly) and smaller doses of water-soluble medications because of their increased body fat and decreased total body water.

There is little evidence suggesting that mild

---

From the Division of General Medicine, Department of Medicine, Mount Sinai School of Medicine, New York, NY 10029. Address reprint requests to the author, who is Chief, Division of General Medicine, at the Department of Medicine, Box 1088, The Mount Sinai Medical Center, One Gustave L. Levy Place, New York, NY 10029.

to moderate obesity increases surgical morbidity or mortality. However, severe obesity increases these risks exponentially to 2–3 times the risk of persons of normal weight. There are multiple height-weight tables and formulae to assess the degree of overweight, but perhaps the most useful bedside tool is the Ponderal Index, arrived at by dividing the height in inches by the cube root of the body weight in pounds. As the Ponderal Index approaches 12.3, operative risk begins to rise, and once it reaches 11.5, morbidity and mortality climb swiftly.

Perioperative problems encountered in the obese include the following

- The major problems center around the respiratory tract. Respiratory depressant drugs used during surgery and anesthesia, the supine position which further limits excursions of the diaphragm, and the physiologic effects of surgery and anesthesia potentiate the defects of obesity and result in increased shunting, ventilation-perfusion mismatches, and hypoxemia. Hypoventilation predisposes to pneumonia, atelectasis, respiratory insufficiency, cardiac arrhythmias and congestive heart failure.
- There is an increased incidence of wound infection, dehiscence, and postoperative incisional hernia in severely obese patients.
- Most general anesthetics are fat soluble so that larger doses may be required for anesthesia and the washout time may be prolonged. Respirations may require more prolonged support postoperatively. Full awakening and extubation are often delayed.
- Although the data are not totally consistent, a consensus does exist that severely obese persons have an increased incidence of postoperative deep vein thrombosis and pulmonary embolism. Such patients have low antithrombin-3 function and concentrations, which in combination with anatomic factors may explain the phenomenon.
- The increased incidence of esophageal reflux places obese patients at higher risk for perioperative aspiration.

Postoperatively, great care must be exercised to provide adequate support for respiration. Incentive spirometry should be taught preoperatively and its practice should be emphasized postoperatively, along with fastidious pulmonary toilet. The patient should be encouraged to sit up rather than to lie down, and should avoid the Trendelenburg position. Early walking is essential. Consideration should be given to minidose heparin therapy.

### Diabetes Mellitus

In the United States there are an estimated 10,000,000 diabetics, 75% of them over age 40. The multiple diseases to which diabetes predisposes, combined with those for which all humankind is at risk, make it a common problem in the preoperative setting. Five points must be emphasized in the evaluation of these patients:

1. **Insulin needs increase intra- and post-operatively. Well-controlled type 2 diabetics may develop a significant insulin requirement during these periods.** Normal individuals secrete 30–40 units of insulin daily. Long-standing, brittle, type 1 diabetics may be able to produce only 5 units/day or less. The stresses engendered by the underlying disease, by surgery, and by anesthesia result in the production of a number of counterregulatory hormones (glucagon, catecholamines, cortisol, and growth hormone), which suppress insulin secretion and cause gluconeogenesis, glycogenolysis, and tissue resistance to insulin. In nondiabetics, the resultant hyperglycemia stimulates insulin secretion, which corrects towards euglycemia and inhibits hepatic ketogenesis. The diabetic, with limited or no reserve, cannot respond with sufficient insulin production. The resultant osmotic diuresis results in fluid and electrolyte depletion, hyperosmolality, hyperviscosity, and thrombogenesis. Increased fatty acid metabolism by the liver eventuates in ketosis and acidosis. These factors are magnified by obesity (decreased insulin-binding sites), steroid ingestion, and infection. Patients undergoing open-heart surgery and organ transplants develop even greater insulin needs (hypothermia, use of insulin counter-reactive drugs, and so on).
2. **Hospital admission of diabetic patients a day or so prior to surgery.** Although many of the required preoperative studies of diabetic patients can be performed as outpatient procedures, diabetics should, whenever possible, be admitted to the hospital a day or so prior to surgery (however much the present climate of DRGs and PROs makes of that aim a non sequitur) to permit:

    **A. Optimization of control of the disease** (blood sugar, pH, volume, osmolality and electrolytes). Utilization of hemoglobin $A1_c$ levels is helpful in assessing recent control. The presence or persistence of blurred vision, polyurea, and polydypsia suggest poor control and the potential need for preoperative fluid and electrolyte correction.

B. **Discontinuation of oral hypoglycemic medications** the night prior to surgery. Chlorpropamide, which has a half-life of 36 hours and the potential for hypoglycemia for 60 hours, should be discontinued 2–3 days preoperatively.

C. **Careful evaluation for target organ involvement,** the major factor in increasing operative risk in diabetics. The increased coincidence of hypertension and diabetes is well recognized, as are the effects of the disease on the micro- and macro-circulation of the heart, the kidney, and the central, autonomic, and peripheral nervous systems, all of which must be thoroughly investigated. There is a high coincidence of coronary atherosclerosis and peripheral vascular disease. Adequate evaluation of angina in such patients may be limited by their inability to exercise; in such situations, the Persantine stress test may be useful. Diuretics used for the control of hypertension or congestive heart failure may complicate surgery by creating volume depletion and hypokalemia.

The presence of chronic renal disease may decrease insulin requirements and often mandates alteration in the choice and dosage of many medications. Recognition of the dangers to diabetically impaired kidneys from volume depletion, nephrotoxic drugs, and intravenous contrast dyes is as important as the search for local infection and the role catheterization may play in it.

When present, autonomic neuropathy may compromise neuroreflex control of the cardiovascular and respiratory responses to stress generated by disease, surgery, and anesthesia. A history of orthostatic hypotension, impotence, or nocturnal diarrhea should alert the physician to this possibility. Autonomic cardiac neuropathy may be reflected clinically by a loss of normal variations in cardiac rate and rhythm with deep breathing.

3. **Early morning surgery should be arranged for diabetics whenever possible.** Surgery later in the day can be compensated for and managed well, but a long period of starvation is not ideal preparation for any diabetic; moreover, if one has chosen to use subcutaneous intermediate-acting insulin the morning of surgery, its midafternoon peak of action may coincide with and further complicate surgery.
4. **Postoperatively, insulin needs may increase significantly, even double, for the first 24–36 hours.** Tighter control is important to promote healing and avoid infection. Careful monitoring of all target organs is mandatory.

   Particular attention must be given to the cardiovascular system; 50% of postoperative myocardial infarctions are painless, and an even larger percentage are painless in diabetic patients. The peak incidence of infarction is between postoperative days 3 and 5. Infarction may be heralded by the onset of an arrhythmia, congestive heart failure, shock, or sudden alteration of mental status. Electrocardiograms should be performed postoperatively and on day 3 and 5.

   Renal function must be checked carefully.

   The majority of postoperative complications in diabetics involve infection, either at the operative site or in the urinary tract. Catheters should be removed at the earliest possible time.

   Gastroparesis, ileus, and urinary retention are common postoperative problems and often are manifestations of autonomic neuropathy.

   In the patient with peripheral vascular disease, careful attention must be given to skin care.
5. **Ideal blood sugar levels in the perioperative period are between 150–250 mg/dL.** Adequate perioperative control is needed to avoid osmotic diuresis, intravascular volume depletion, electrolyte disturbances, acidosis, and hyperosmolality and to maintain defenses against infection. Equally critical is the avoidance of hypoglycemia, which can be catastrophic in a narcotized or anesthetized patient, particularly if the patient has some underlying vascular disease.

**Perioperative Diabetic Management.** Central to each of the multiple approaches to the perioperative management of diabetics is the physician's comfort and experience with the method used and close observation of the patient, including monitoring of the blood sugar every 2 hr. The following is a synthesis of guidelines to the perioperative management of diabetics. No one protocol can cover every situation, and clinical judgment always take precedence.

**Well-controlled diabetics undergoing early-morning minor surgery:** hypoglycemic agents may be withheld and pre-, intra-, and post-operative blood sugars are obtained. If the postoperative blood sugar is less than 250 mg/dL and the patient is able to eat lunch, oral hypoglycemics are then resumed or, in patients with insulin-dependent diabetes mellitus, 50% of the usual dose of intermediate-acting insulin is administered subcutaneously. If the blood sugar is greater than 250 mg/dL, subcutaneous regular insulin may be administered on a sliding scale every 4–6 hours for the remainder of the day and the usual regimen resumed the following morning.

**Diet-controlled diabetics undergoing major surgery:** intraoperative blood sugars should be obtained and if levels exceed 250 mg/dL in this or the post-operative period, human subcutaneous regular insulin (5–10 units) or intravenous insulin (1 unit/hr) may be given with appropriate glucose and electrolyte supplementation.

Some nonobese uninfected patients who have been well controlled on oral hypoglycemic agents may be managed as above during major surgery. More often, they have insufficient pancreatic reserve to meet the increased demands and will require some insulin in the perioperative period. These needs may be met via the subcutaneous or intravenous route, but many clinicians prefer to begin treatment of these patients preoperatively with a constant infusion of insulin (0.5–1.0 unit/hr) along with 100 mL 5% dextrose and half-normal saline hourly, modifying the insulin dose as dictated by frequent intra- and post-operative blood sugar determinations.

**All insulin-dependent diabetics undergoing emergency or elective major surgery:** insulin coverage is required. This may be accomplished by the subcutaneous regimen or by one of two continuous low-dose insulin infusion methods. Common to all is the need for frequent monitoring of the blood sugar to avoid extremes of glycemia. Such factors as diabetic control, brittleness, tendency to acidosis, the presence of infection or obesity, the use of exogenous steroids, a team who are experienced in and comfortable with the method selected, and the nature of the surgery may play a role in the choice of regimens. The continuous low-dose insulin infusion method is becoming more popular because of its flexibility and improved moment-to-moment control.

**Insulin Regimens.** In the subcutaneous method, 50% of the usual dose of intermediate acting insulin is given preoperatively and an intravenous infusion of 5% dextrose and half-normal saline is begun and run at a rate of 100–150 ml/hr throughout the procedure. Fluctuations in glucose levels are managed by altering the rate of the infusion, increasing the concentration of glucose, or by adding insulin (subcutaneous or intravenous). Postoperatively, if the blood sugar is between 150–250 mg/dL, the intravenous infusion is continued and the other half-dose is given subcutaneously.

The low-dose continuous infusion method involves the administration of 1–3 units of regular insulin along with 100 mL of 5% dextrose and half-normal saline per hour. Multiple protocols are available and should be consulted. Although the hour-to-hour dose will be mandated by the blood sugar levels, some guidelines are useful: if the usual daily insulin requirement was less than 40 units, start with 1 unit/hr; if the daily need was 40–80 units, start with 1.5 units/hr; and if the daily dose was 40–80 units, start with 2 units/hr. If the glucose dips below 120 ml/dL, decrease the insulin by 0.5 units/hr; if it rises above 250–300 mg/dL, increase by 0.5 units/hr; and if it rises above 300–350 mg/dL, increase by another 0.5 units/hr. Higher levels may require an insulin bolus of 5–10 units, lower levels are managed by discontinuing the insulin and, if necessary, by administering 50% dextrose. Since insulin binds to infusion equipment and glassware, it is worthwhile to first flush the unit, discarding the first 50 mL or so.

The insulin may be given intravenously by diluting 10, 15, or 20 units of regular insulin in 1000 mL of 5% dextrose and half-normal saline (yielding concentrations of 1, 1.5, or 2 units of insulin per 100 mL). This solution then is given at a rate of 100 mL/hr. As the blood sugar varies throughout surgery, the intravenous bottles must be changed, the rate of infusion (and thus of fluid replacement) altered, or insulin added to the bottle to meet the changing requirements. A more efficient and simpler practice is to make a solution of 250 units of insulin in 500 mL of 5% dextrose (0.5 units/mL) and, using an insulin pump, piggyback the infusion, permitting delivery of the proper amount of insulin independent of the rate or contents of the IV. Postoperative management may be via 4–6 hr subcutaneous insulin on a sliding scale or via continuous low dose infusion. Potassium replacement is usually required (20–40 mEq of KCl/L). The patient returns to the usual insulin dose when able to eat.

**Diabetic Ketoacidosis.** The patient coming to surgery with diabetic ketoacidosis represents an operative risk sufficiently high to mandate the delay of surgery in all but the most life-threatening situations. The short time (4–6 hours) required to sufficiently correct these metabolic abnormalities and replete intravascular volume also permits rapid evaluation and therapy of any target organ involvement or infection. Space does not allow a recapitulation of the various protocols for the management of this condition, but several points should be emphasized.

- Ketoacidosis can mimic an acute abdomen. It may be accompanied by abdominal pain (which subsides as the acidosis is controlled) and ileus. In these patients, the level of serum amylase is often elevated.

- These patients have severe fluid deficits (averaging 5 L but can require as much as 0.1 L of fluid/kg body weight for full correction). The 3–4 L of fluid given in the first 4–6 hours of management may not completely replace deficits. Underlying cardiac or renal disease will influence, and may limit, the speed, adequacy, or content of acute replacement.
- These patients are salt depleted. In evaluating serum sodium levels, recall that every 100 mg/dL increase in glucose results in a 1.6 mEq/L decrease in sodium as a result of osmotic mechanisms (intracellular movement), and that hyperlipidemia and hyperproteinemia result in spuriously low serum sodium determinations.
- These patients may have a total body potassium deficit of 300 mEq prior to therapy, and as the metabolic defect is corrected and potassium moves intracellularly, serum levels decrease and potassium must be replaced via the intravenous route. These needs may continue intraoperatively and must be monitored and met.
- As the blood sugar approaches 250–300 mg/dL, a shift in fluids to 5% dextrose and half-normal saline should be made to protect against hypoglycemia. This must be continued intraoperatively and later, until the patient is able to tolerate oral feedings.
- Hypernatremia, dehydration, and hyperglycemia in the absence of acidosis should suggest a hyperosmolar state requiring larger amounts of half-normal saline and lesser amounts of insulin. This state is more common in the elderly patient taking oral hypoglycemic drugs.

Under the best of circumstances, the diabetic treated for ketoacidosis for 4 or so hours enters the operating room a better risk than prior to therapy albeit with some remaining fluid and electrolyte deficit. Since the occurrence of hypoglycemia is still possible, careful monitoring is essential. If significant renal or cardiac disease is present or a real question about fluid status exists, the use of invasive monitoring should be considered.

## The Adrenal Glands

**Adrenal Insufficiency.** The adrenal gland of a healthy adult secretes 15–20 mg of cortisol daily; under stress it may produce up to 200–500 mg. During major surgery, cortisol levels rise from a preoperative baseline of ±10 μg/dL to 17–18 μg/dL during the first half hour after induction and may peak at 50 μg/dL. Postoperatively, levels may remain near 25 μg/dL and not attain near-baseline values for 3 days, remaining slightly elevated for up to 7–10 days. During minor surgery, they approach 30 μg/dL, returning to near basal levels in 24 hours. In patients who have had suppressive doses of cortisol in the previous year, baseline levels may be much the same as in non-adrenal-insufficient persons, but there may be little or no rise at the half-hour, and at an hour after induction, levels often reach only 50%–60% of normal values. *Once suppressed for a period as short as a week by exogenous steroids, normal adrenal responsiveness may not return for a year. The stresses of surgery and anesthesia can uncover and convert this relative adrenal insufficiency to frank hypoadrenalism.* Patients who have sustained endogenous suppression by tumor or hyperplasia may require up to two years to regain normal responsiveness because of the prolonged period of suppression.

A careful history, searching not only for cortisol usage but also for the type, dosage, and chronicity of such therapy, is essential, particularly in patients with diseases for which the drug might be prescribed (arthritis, asthma, chronic obstructive pulmonary disease, autoimmune and allergic conditions, hematologic and oncologic problems, inflammatory bowel disease, among others).

**In general, the danger of suppression is less with:**

- shorter-acting adrenal steroids (cortisone, hydrocortisone, prednisone, and so on, with effects lasting 24–36 hours) than with longer-acting ones (dexamethasone and betamethasone, with effects lasting over 48 hours).
- once-a-day (early AM) doses rather than divided doses. Once-daily every-other-day therapy with short-acting steroids in doses under 80 mg is almost never associated with adrenal suppression. Occasionally, baseline cortisol levels will be low in such patients, but the ability to respond adequately to stress is preserved.
- steroid inhalants and skin preparations, or short courses of ACTH; neither treatment is likely to suppress the patient's normal responsiveness.

The long-acting agents have little mineralocorticoid activity, whereas 25 mg of hydrocortisone a day has sufficient activity for such replacement.

**Suppression is likely and coverage prudent** in those who in the past year have had over 5–7 days of treatment with:

- a long-acting preparation
- 20 mg of prednisone or supraphysiologic doses of any steroid
- any adrenal steroid taken in divided doses. Prednisone taken in doses of 5–7.5 mg/day for over a month is much less likely to result in adrenal suppression. However, these patients may have a reduced response of the hypothalamic-pituitary-adrenal axis to induced hypoglycemia, making coverage prudent.

Time does not usually permit awaiting the results of the rapid ACTH stimulation test, which would confirm the presence or absence of adrenal suppression. If baseline cortisol levels are over 22 μg/dL, adrenal suppression is unlikely. In normals, the intravenous administration of 25 units of synthetic ACTH is followed by a rise in cortisol levels in an hour to 18–20 μg/dL, with a doubling of the basal levels and an absolute rise of 7 μg/dL. *If there is a question, treat.* The potential problems from such short-term therapy of altered or delayed healing or infection, hypokalemia, hyperglycemia, and fluid retention are less than the potential problems of perioperative adrenal insufficiency.

Patients with adrenal insufficiency because of destruction of the adrenal glands have an added mineralocorticoid defect. Usually, there is sufficient mineralocorticoid activity in the stress doses of hydrocortisone administered to cover these needs, but coverage must be monitored, and, on occasion, added saline or small doses of Florinef are required.

For major surgery, 100 mg of hydrocortisone hemisuccinate is given intravenously 1 hr preoperatively and repeated every 6 hr on the operative and first postoperative days. If no problems arise and the patient does well, the dose is cut in half on the 2nd postoperative day and again halved the following day. Thereafter, it is decreased to maintenance levels. Minor procedures (herniorrhaphy for example) are handled similarly, but the steroids are cut back to maintenance more rapidly (usually in 2 days). For procedures such as endoscopy and arteriography, 100 mg 1 hr prior to the procedure usually suffices. Dexamethasone in doses of 4 mg can be given in the same manner, preserving the validity of the ACTH stimulation test, thus facilitating discontinuation of the steroids postoperatively in the nonsuppressed. The lack of mineralocorticoid activity of dexamethasone is usually compensated for by the administered saline.

**Pheochromocytoma.** In addition to the risks of hypertension and hypertensive crisis and their sequelae, patients with pheochromocytoma are generally volume depleted and predisposed to tachyarrhythmias and orthostatic hypotension. These patients are exquisitely sensitive to all catecholamine-releasing drugs. Whether the patient is being operated on for the pheochromocytoma or for another condition, or is simply having an intraarterial radiographic study with contrast dye, proper preparation is essential.

**Preoperative preparation** should begin two weeks prior to proposed surgery. Such therapy is anchored in the management of hypertension and cyclic symptoms with alpha blockers and repletion of the shrunken plasma volume. As the catecholamine levels are reduced by therapy, the extracellular volume increases. In patients in whom there is no contraindication, this is facilitated by the addition of liberal amounts of salt to the diet. As a result, a drop in the hematocrit is not unusual during this period.

Phenoxybenzamine (Dibenzyline), a noncompetitive alpha blocker, is the drug of choice. The peak effect of a given dose is seen in 12 hr; it has cumulative effects over several days which may persist for 3–4 days. Therapy is begun in doses of 10 mg twice daily and increased by this amount every 2–3 days until the blood pressure and paroxysmal symptoms are controlled. Usually 40–80 mg/day suffices, but occasionally as much as 200 mg/day is required. The drug does not completely block the alpha receptors, so that if this is the sole therapy utilized, some risk of hypertensive crisis remains during periods of maximal sympathetic stimulation. Some protocols utilize this drug up to the time of surgery and manage intraoperative hypertension with nitroprusside or phentolamine by intravenous bolus or drip. Others prefer to hospitalize these patients 48 hr prior to surgery, discontinue the phenoxybenzamine, and stabilize the patient pre- and intraoperatively on phentolamine (Regitine), an alpha blocker with a rapid onset and short duration of action, and, at the same time, take the opportunity to fully replete the intravascular volume.

Demser R (alpha-methyl paratyrosine) can inhibit the synthesis of catecholamines by 40%–80%. Alone, it seldom completely controls the blood pressure, but does decrease the dose of alpha blocker necessary to do so. In high doses, crystalluria and extrapyramidal side effects are seen. Its major use is in the management of malignant pheochromocytomas.

Beta blockers should not be used without prior alpha blockade. Such therapy results in unopposed alpha-adrenergic stimulation and the risk of paradoxic hypertensive crises. Inderal, when used along with alpha blockade, is effective in management of the tachycardia and supraventricular arrhythmias associated with this disease.

**At surgery** these patients require close monitoring. They require ECG, intraarterial and central lines, and prudence may dictate a Swan-Ganz catheter as well. Preanesthetic and anesthetic agents should be selected from among drugs with the least ability to stimulate catecholamine release or sensitize their receptors.

Intraoperatively, extracellular volume must be maintained. Hypertension is managed with intravenous phentolamine nitroglycerine or nitroprusside. Induction, intubation, and manipulation of the tumor are the periods of highest risk for hypertensive responses. Hypotension is best handled by volume expansion (whole blood sufficient to replace loss, saline, and plasma expanders). There may be relative refractoriness to infused catecholamines. Supraventricular arrhythmias respond well to Inderal, those of ventricular origin to lidocaine. Electrical equipment should be available for management of nonresponsive arrhythmias.

**Postoperatively**, transient hypertension resulting from fluid shifts may occur; it usually responds to diuretics. Catecholamine levels may not return completely to normal for up to a week or two, and some hypertension may persist during this period. More protracted blood pressure elevations raise the question of remaining tumor or accidental occlusion of the renal artery.

**Hyperaldosteronism.** Perioperative problems arising in patients with hyperaldosteronism center around correcting the hypokalemia and maintaining control of the blood pressure. The potassium deficit may be large, even in the presence of moderate decreases in serum levels.

**Cushing's Syndrome.** The hyperglycemia, electrolyte abnormalities, obesity and hypertension associated with Cushing's syndrome should be managed preoperatively. The prolonged hypercorticism results in more prolonged suppression of responsiveness of the axis and may contribute to delayed wound healing and infection postoperatively. When time permits, consideration should be given to preoperative use of inhibitors of steroid synthesis to minimize these risks. These patients require postoperative steroid supplementation.

### Thyroid Disease

The thyroid gland, the only source of $T_4$, normally secretes 90–100 μg of $T_4$ and 5 μg of $T_3$ daily. Deiodination, the major route of $T_4$ disposal, is under tight, although poorly understood, metabolic control. Of the 30 μg of $T_3$ (the active hormone) generated each day, 25 μg arises from the 5′ deiodination of $T_4$ by the peripheral tissues; 99% of the metabolically inert reverse $T_3$ is generated by deiodinations at the 5 position. On release from the thyroid gland, $T_3$ and $T_4$ are immediately and almost completely bound to specific plasma proteins in a firm but reversible bond with only minute amounts existing in the free, metabolically active state (0.02%–0.04% of $T_4$ and 0.2%–0.4% of $T_3$). $T_3$ and $T_4$, as determined by radioimmunoassay, measure total hormone (free and bound). Free hormone remains in equilibrium with that bound to plasma proteins in such a manner that alteration in the level of these proteins, although affecting the measured levels of total $T_3$ and $T_4$, does not alter the level of the free hormone or the metabolic state. In normals, hormonal homestasis is maintained by a sensitive servomechanism; increased levels of free hormone result in suppression of thyroid-stimulating hormone (TSH) production and then decreased thyroidal production; a fall in the circulating levels of $T_3$ and $T_4$ results in the converse.

Since thyroid hormones are >99% protein bound, diseases (anabolic states, cachexia) and medications (anabolic steroids) that decrease these proteins will lower the serum concentration of the thyroid hormones, and those that increase these proteins (estrogens, liver disease) will raise them. It is the small amount of free hormone which determines the metabolic state, and since this is not altered by changes in the binding proteins, the serum levels of total $T_3$ and $T_4$ will not reflect the metabolic state in such situations. Binding protein levels ($T_3$ resin uptake, or TBG) or free hormone determinations are also needed for proper interpretation.

Most acute and chronic illness, caloric deprivation, surgery, anesthesia, and a multiplicity of drugs (cortisol, beta blockers, iodinated dyes, amiodarone) can inhibit the 5′ deiodination of $T_4$ to $T_3$, rendering values for these moieties that do not reflect the metabolic state of the patient (the eumetabolic low $T_3$, or sick euthyroid syndrome).

Acute and chronic illness may suppress TSH secretion or, if there is sufficient inhibition of $T_4$ deiodination, may result in a slight rise. Dopamine and cortisol suppress TSH levels and the pituitary response to TRH. This may also be seen in DSM-III depressions and in the aged, particularly men.

Hyperthyroidism may be masked by medications (beta blockers) and may have an atypical initial clinical appearance in the aged. Both latent hyper- and hypo-thyroidism may be uncovered by acute illness, or the presenting disease may be so prominent as to mask an underlying thyroid condition.

An appreciation and understanding of the physiologic mechanisms involved will simplify many of the consultations requested by the surgical services, since such requests are often based on abnormalities in thyroid testing results, not thyroid function.

The most effective screening for hyperthyroidism is $T_3$ by radioimmunoassay, a measure of binding proteins, and a TSH level; for hypothyroidism, the most effective screening is a $T_4$, TSH, and TBG or $T_3$ resin uptake. Since so many drugs and diseases can result in a low $T_3$, the test is valueless in the diagnosis of hypothyroidism. In general, $T_4$ hyperthyroidism is rare, and return of an elevated $T_3$ to normal via inhibition of $T_4$ deiodination by drugs or disease is equally unusual. Suppression of significantly elevated TSH levels to normal in a clinically hypothyroid patient is equally unlikely. Given one test, except in the rare instances of central disease, the supersensitive TSH is almost invariably low in hyperthyroidism and high in hypothyroidism.

**Hyperthyroidism.** Patients with hyperthyroidism that is well controlled on antithyroid drugs do not present any increased operative risk. The residence time of these agents in the thyroid gland is over 24 hr, permitting discontinuation of the drug for several days without ill effect. Coverage with iodides and beta blockers can be undertaken if any question exists.

Untreated or undertreated hyperthyroid patients present a real risk of thyroid storm and, with it, high mortality. Whenever possible, surgery should be deferred until the patient has been rendered eumetabolic. If the urgency of the situation does not permit this, the following preoperative protocol may be utilized:

- Identify and treat any underlying infection or precipitating event.
- Give a loading dose of an antithyroid drug. Propylthiouracil (400 mg) has the advantage of inhibiting $T_4$ deiodination and is preferred. At this juncture, its only purpose is to inhibit organification of administered iodides and thus prevent enrichment of the thyroid gland with this element (facilitating whatever long-term therapy is decided upon).
- Add 1 g of sodium iodide to the IV twice daily to block thyroidal release of hormone.
- Inderal, orally or parenterally, should be started to block the sympathetomimetic effects of $T_3$ and $T_4$. Both Inderal and cortisone also block $T_4$ deiodination to $T_3$.
- Give stress doses of cortisone—these patients may have relative adrenal insufficiency (rapid turnover of cortisol may result in the pituitary adrenal axis functioning at such a high basal level as to render it unable to further respond to added stresses).
- Sodium ipodate (Orografin, an organic gallbladder dye) in doses of 500 mg–1 g/day almost completely inhibits deiodination of $T_4$ (in 24 hr it may reduce the levels of $T_3$ by 50%).

Postoperatively, this protocol is continued and gradually altered to meet the patient's needs. As oral intake is resumed, the PTU should be restarted in full therapeutic doses.

**Hypothyroidism.** The metabolic defects of hypothyroidism are multiple, and many are synergistic with those induced by anesthesia and surgery:

- decreased cardiac output and stroke volume, bradycardia, atrioventricular conduction defects, and pericardial effusion
- peripheral vasoconstriction
- decreased alveolar ventilation due to lessened central nervous system responsiveness to anoxia and hypercarbia, respiratory muscle myopathy, interstitial infiltration with mucopolysaccharide and pleural effusions
- decreased renal blood flow resulting in decreased glomerular filtration rate, with less water reaching the distal collecting system and resultant increased reabsorption (this largely explains the inability to handle a water load and can eventuate in hyponatremia
- altered iron, vitamin $B_{12}$, and folate metabolism, with resultant anemia
- deficiency of 2,3-DPG eventuates in shifting the $O_2$ dissociation curve to the right
- decreased rate of metabolism of administered drugs; a sensitivity to narcotics and analgesics coupled with respiratory deficits sets the stage for respiratory failure
- altered mentation
- tendency to ileus.

Hypothyroidism is associated with elevations of the CPK, LDH, SGOT, uric acid, and blood lipids.

Patients with well-controlled hypothyroidism do not present any enhanced operative risk. The half-life of $T_4$ in the blood is 7+ days and a large reservoir exists attached to the plasma proteins. Replacement therapy can be discontinued for a week or so without any problem.

Whenever possible, surgery should be deferred until hypothyroidism is corrected. Several small studies suggest that patients with mild to moderate hypothyroidism can be operated on, uncorrected, with little increase in cardiac arrhythmias, hypothermia, hyponatremia, delayed recovery from anesthesia, or increase in pulmonary complications. There was an increased incidence of intraoperative hypotension and postoperative congestive heart failure, confusion, and ileus. These studies suggest that intravenous thyroid hormone is not necessary in such situations but do not negate the need for careful monitoring of respiratory status, limitation of sedation and narcotics postoperatively, and wariness in regard to fluid overload. Despite these studies, many consultants will recommend preoperative oral administration of small doses of thyroid hormone, if not otherwise contraindicated.

With the exception of patients undergoing cardiac bypass surgery (who do quite well without any replacement and in whom thyroid hormones often increase angina and arrhythmias), patients with severe hypothyroidism usually require some hormone replacement preoperatively. Depending on the age of the patient, the severity of the hypothyroidism, and associated medical conditions, 300–750 μg of intravenous $T_4$ is given slowly. Stress doses of cortisol are given to protect against a sluggish adrenal response to the added demands of anesthesia, surgery, the condition precipitating surgery, and administered thyroid hormone. Careful cardiopulmonary monitoring is essential, premature extubation must be avoided, and limited use of sedation and narcotics is essential, as is meticulous attention to fluid intake to prevent overhydration and hyponatremia. An altered mental state and mild respiratory difficulty may be the earliest signs of impending myxedema coma. The joint use of thyroid hormone and pressor agents increases the incidence of cardiac arrhythmias; 50–100 μg of $T_4$ may be given daily by the intravenous route until the patient is able to tolerate oral feedings.

## Other Endocrinopathies

**Hypopituitarism.** Established hypopituitary patients scheduled for surgery must receive stress doses of cortisol in the perioperative period. When able to resume oral feedings, they may be restarted on the remainder of their replacement therapy. If vasopressin deficiency exists, great care must be taken to replace early and adequately, as the loss of thirst perception in the narcotized state can result in rapid, serious dehydration.

Most acute hypopituitarism seen on the surgical services follows head trauma or surgical ablation. Pituitary apoplexy occurs post partum, as well as with diabetes, vasculitis, sickle-cell crisis, and bleeding into a pituitary tumor. The earliest endocrinologic manifestation is hypoadrenalism (24–36 hr after surgery). The cornerstones of therapy are volume expansion and cortisol. Insulin needs decrease in hypopituitarism.

**Diabetes Insipidus.** Diabetes insipidus following pituitary injury may be immediate or delayed and is often transient (starting on the 1st–3rd postoperative day and lasting about a week). Although definitive diagnosis may require a fluid deprivation test, generally patients with this form of diabetes have urine specific gravities below 1.005, with osmolalities of 50–200 mOsm/kg (<300) in the presence of hypernatremia and increased serum osmolality. Nephrogenic diabetes insipidus may occur as a result of hypokalemia or hypercalcemia; the ingestion of lithium, demeclocycline, or amphotericin; or enthrane or methoxyflurane anesthesia.

Therapy consists of dextrose in water and aqueous pitressin (5–10 units, subcutaneous, every 4–6 hr) or DDAVP (desmopressin, 1–2 μg intravenous or subcutaneous every 12 hr). The hypernatremia and hyperosmolality of excessive sodium administration (as from sodium bicarbonate therapy in severe metabolic acidosis) or of increased free water loss with inadequate replacement (as in burns, gastrointestinal losses, vigorous osmotic diuresis as from mannitol, glucose, or urea) are associated with high urine osmolalities (greater than 600). Iatrogenic fluid overload followed by osmotic diuresis is characterized by hyponatremia and low normal or slightly decreased serum osmolality with urinary specific gravities over 1.010 and osmolalities of 250–300.

**SIADH.** The administration of excessive quantities of hypotonic solutions, not SIADH, is the most common cause of postoperative hyponatremia. In addition, the postoperative state is associated with increased endogenous water production (increased oxidation of the catabolic state). These factors, together with the continued appropriate secretion of ADH, are responsible for the postoperative hyponatremia in patients with normal renal function. Of the multiple stimuli (anesthesia, pain, narcotics, mechanical ventilation, hypoxia, and hypercarbia) potentiating the secretion of ADH in the perioperative period, volume depletion is the one most often responsible.

In the absence of these stimuli, and when hypothyroidism, hypoadrenalism, congestive heart failure, and hepatic failure have been excluded, the likelihood of a diagnosis of SIADH in a hyponatremic patient is greater when the urinary excretion of sodium exceeds 20 mEq/L, the urine is less than maximally dilute (>200 mOsm/kg), and the plasma is hypotonic (<285 mOsm/kg). In the surgical setting, the condition is most often seen on the neurosurgical services (head trauma, subarachnoid hemorrhage, tumor), in patients with ADH-secreting tumors (pancreas, lung, and so on) and those with pneumonia. Thiazides, tricyclic antidepressants, phenothiazines, barbiturates, chlorpropamide, vincristine, and cyclophosphamide can cause the syndrome as well. Fluid restriction is usually sufficient therapy. For symptomatic hyponatremia (sodium less than 110–115), hypertonic saline and furosemide may be required. Declomycin (300 mg 2–4 times a day) may also be utilized.

Although direct measurement of serum and urinary osmolalities is usually quickly available, a useful formula for calculating serum osmolality within 5% from commonly available blood chemistry levels is:

$$\text{Serum osmolality} = 2(\text{Na} + \text{K}) + \text{blood sugar}/18 + \text{BUN}/28.$$

**Hypercalcemia.** The causes of hypercalcemia are multiple and must be identified preoperatively; in the hospital setting, malignancy is most common. Acute increases in serum calcium are associated with myocardial effects (decreased automaticity and increased contractility manifest by heart block, arrhythmias, and potentiation of the effects of digitalis), altered mental state, constipation, and ileus. Chronic hypercalcemia may be accompanied by an increased incidence of peptic ulcer disease, kidney stones, and nephropathy. In chronically ill patients every gram alteration in serum albumin is accompanied by an 0.2 μM/L change in calcium levels in the same direction. Permitting a tourniquet to remain in place for several minutes can result in a spurious elevation of calcium of 0.125–0.25 μM/L.

Serum calcium levels should be reduced to 2.75–2.8 μM/L prior to surgery. Frequently this can be accomplished by a combination of saline infusions (1–2 L 4 times a day), furosemide, and cortisol. Since hypovolemia leads to increased calcium absorption, a brisk diuresis should be obtained with normal saline prior to the administration of 40–80 mg of intravenous furosemide as needed. Together with cortisol, which prevents calcium absorption from the gut, this combination is usually sufficient to reduce calcium levels. Such therapy may result in hypokalemia and hypomagnesemia, and these electrolytes must be monitored and replaced as indicated. Calcitonin, 4–8 MRC units per kg given every 6–12 hr, may result in a reduction in serum calcium in 1–2 hr, a reduction prolonged by the use of glucocorticoids. Recent experience with combinations of calcitonin and diphosphonates is promising in the management of more resistant cases. Plicamycin (mithromycin) can reduce calcium acutely and transiently, but its toxicity in marrow (early thrombocytopenia), liver (inhibition of hepatic clotting factors), and kidney limits its usefulness preoperatively and in nonmalignant disease. Intravenous infusion of inorganic phosphate solutions containing 1.5 g elemental phosphorus given over 6–8 hr is effective in minutes, but can result in metastatic calcification. The usefulness of EDTA is limited by the renal toxicity caused by EDTA-calcium complexes. Unresponsive hypercalcemia may be managed by dialysis.

**Estrogens.** Estrogens can prolong the central nervous system depressive effects of many analgesic and anesthetic agents, most likely as a result of decreased hepatic metabolism. In premenopausal women, estrogens increase the incidence of postoperative thromboembolism almost fivefold. This is a dose-related phenomenon and does not appear to hold in the postmenopausal state. When possible, these medications should be discontinued 3–4 weeks prior to surgery; if that is impossible, minidose heparin should be considered.

## Suggested Reading

Strauss R, Wise L. Operative risks of obesity. Surg Obstet Gynecol 1978; 146:686.

Pesulka PS, et al. The risks of surgery in obese patients. Ann Intern Med 1986; 104:540.

Seltzer CC. Some reevaluations of the build and blood pressure as related to ponderal index, somatotype and mortality. N Engl J Med 1966; 274:254.

Chase HF. To operate or not to operate—what is the risk? Res Staff Phys 1978; 24:65.

Kakkar VV, et al. Deep vein thrombosis of the leg. Is there a high risk group? Am J Surg 1970; 120:527.

McMurry J. Wound healing with diabetes mellitus. Surg Clin North Am 1984; 64:769.

Alberti KGMM, Thomas DJB. The management of diabetes during surgery. Br J Anaesth 1979; 51:693.

Podolsky S. Management of diabetes in the surgical patient. Med Clin North Am 1982; 66:1361.

Taitelman V, et al. Insulin in the management of the diabetic surgical patient. Continuous intravenous infusion vs subcutaneous administrations. JAMA 1977; 237:658.

Reynolds C. Management of the diabetic surgical patient. A systematic but flexible plan. Postgrad Med 1985; 77:265.

Galloway JA, Schuman CR. Diabetes and surgery; a study of 667 cases. Am J Med 1963; 34:177.

Oyama T. Hazards of steroids in association with anesthesia. Can Anaesth Soc J 1969; 16:361.

Plumpton FS, et al. Corticosteroid therapy and surgery. Anesthesia 1969; 24:12.

Streck WF, et al. Pituitary-adrenal recovery following short term suppression with corticosteroids. Am J Med 1979; 66:910.

Graber AL, et al. Natural history of pituitary-adrenal recovery following long term suppression with corticosteroids. J Clin Endocrinol Metab 1965; 25:11.

Byyny RL. Preventing adrenal insufficiency during surgery. Postgrad Med 1980; 67:219.

Ackerman GL, Nolan CM. Adrenocortical responsiveness after alternate day corticosteroid therapy. N Engl J Med 1968; 278:405.

Gilbert P. Thyroid disease in the elderly. In: Geriatric medicine annual 1986. Oradell, NJ: Medical Economics Books, 1986:213–237.

Wu SY, et al. Comparison of Na ipodate (Orografin) and propylthiouracil in the early treatment of hyperthyroidism. J Clin Endocrinol Metab 1983; 54:630.

See DM, et al. Preoperative endocrine and metabolic considerations. Med Clin North Am 1978; 63:1321.

Landerson PW, et al. Complications of surgery in hypothyroid patients. Am J Med 1984; 77:261.

Weinberg AD, et al. Outcome of anesthesia and surgery in hypothyroid patients. Arch Intern Med 1983; 143:893.

Lukomsky GI, et al. Preoperative treatment of patients with thyrotoxicosis. Am J Surg 1982; 147:766.

Maguire M, et al. Increased risk of thrombus due to oral contraceptives. Am J Epidemiol 1979; 110:188.

Notelovitz M, et al. Combination estrogen and progesterone replacement therapy does not adversely affect coagulation. Obstet Gynecol 1984; 62:596.

# Preoperative Evaluation of the Patient with Chronic Renal Disease

PAUL L. GILBERT, M.D., AND RICHARD STEIN, M.D.

THE preoperative medical evaluation of patients with advanced, chronic renal disease can present an interesting challenge for the consulting internist. The decreased ability of diseased kidneys to regulate and maintain appropriate fluid and electrolyte balance, to excrete metabolic wastes, and to maintain normal body pH is often compounded by underlying heart disease, hypertension, and the associated anemia, malnutrition, coagulopathies, impaired healing, and altered immunocompetence attendant on the uremic state.

The Goldman Cardiac Risk Index recognizes blood urea nitrogen (BUN) levels over 50 mg/dL and creatinine levels over 3 mg/dL as sufficient to increase surgical risk. Once potentially reversible disease is excluded, the etiology of the underlying renal disease is prognostically less important than is the level of renal function.

- As a rule, surgery is well tolerated in patients whose glomerular filtration rate (GFR) remains over 25–30 mL/min (25% of normal). These patients generally have only minimal alterations of acid-base, fluid, and electrolyte balance and are without uremic symptoms. If significant alterations in these modalities do exist at this level of renal function, extrarenal causes should be sought and treated.
- As the GFR approaches 10–15 mL/min, surgical morbidity rises rapidly, approaching 55%–60%, and preoperative dialysis should be seriously considered.

When time does not permit the luxury of clearance studies, an accurate bedside estimate of the GFR can be determined by the following formula, which, in part, corrects for the changes in renal function that occur as part of the normal aging process:

$$\text{GFR} = (140 - \text{age}) \times (\text{weight in kg}) \div (72) \times (\text{serum creatinine in mg/dL}) \; (\times\ 0.85 \text{ for women})$$

## Effects of Anesthesia and Surgery on Renal Function

**Effects of Anesthesia.** Almost all general anesthesia results in decreased renal blood flow (up to 50%), liberation of multiple vasopressor substances, ADH, and activation of the renin-angiotensin system. Anesthesia impairs renal autoregulation, which may be further diminished by intraoperative hypovolemia or hypotension.

Cholinesterase levels are low in chronic renal disease, and prolonged respiratory paralysis may be produced by any of the neuromuscular blocking agents. The fluorinated inhalation agents methoxyflurane and enflurane are potentially nephrotoxic and should be avoided in such patients.

Regional anesthesia offers a greater margin of safety under certain circumstances. The coagulation profile (especially bleeding time) should be normalized preoperatively and renal perfusion well maintained intraoperatively.

**Effects of Surgery.** Nausea, vomiting, diarrhea, nasogastric suction, third-space effects of disease and surgery, physiologic effects of the underlying surgical condition, extracellular fluid volume depletion, hypotension, hypotonic fluid replacement in the face of isotonic fluid and electrolyte losses, aldosterone effects, and hyperali-

---

From the Divisions of General Medicine and Nephrology, Department of Medicine, Mount Sinai School of Medicine, New York, NY 10029. Address reprint requests to the second author, who is Chief, Division of Nephrology, at the Department of Medicine, Box 1088, The Mount Sinai Medical Center, One Gustave L. Levy Place, New York, NY 10029.

mentation all have an impact on fluid and electrolyte balance in the perioperative period. Acidosis, tissue injury, rhabdomyolysis, succinylcholine, and hemolyzed blood (via transfusion, absorption from the gastrointestinal tract, and so on) potentiate hyperkalemia.

Patients with normal kidney function are generally able to cope with and compensate for these alterations in physiology. This ability may become impaired in those with chronic renal insufficiency when GFR falls below 25 mL/min. The risk of clinically significant postoperative renal failure is less than 1% in the group with normal renal function and over 50% in those with low GFRs (up to 90% in high-risk operations). Consequently, preoperative prophylactic dialysis in patients with tenuous renal function (GFRs of 10–15 mL/min) may permit the surgical patient to better tolerate the multiple added physiologic insults of anesthesia and surgery. Moreover, in these patients, the operative procedure may exacerbate the underlying renal insufficiency, making it more likely that dialysis will also be required in the immediate postoperative period. The greater the preoperative problems with bleeding, volume overload, hyperkalemia, and acidosis, particularly in the presence of sepsis, gastrointestinal bleeding, and congestive heart failure, the more sensitive the clinician must be to the need for pre- and/or postoperative dialysis.

Dialysis itself, for whatever indication, can cause problems with hemostasis. The dialysis is often followed by a transient drop in the platelet count, although platelet function itself does improve.

Heparin, even when used sparingly in dialysis, can cause intra- and post-operative bleeding. Regional heparinization (addition of protamine into the infusion distal to the coil) has been employed at some centers. However, low-dose heparin or dialysis without heparin is generally preferred.

If urgent preoperative dialysis is required, surgery should be deferred for 24 hr to minimize the bleeding risks. If 24-hr deferral is impossible, at least 2–6 hr should elapse. Patients being treated with long-term dialysis should have any preoperative transfusion early in their final dialysis some 24 hr prior to surgery to permit the anticoagulant effects to dissipate and to permit correction of fluid balance, pH, and serum potassium levels.

## Risk Factors for Postoperative Renal Failure

Factors that increase the risk of the patient developing postoperative renal failure include the following nine characteristics:

- severity of underlying renal disease
- complexity of surgery performed
- extracellular fluid volume depletion
- hypotension
- sepsis
- exposure to nephrotoxic agents (aminoglycosides, iodinated contrast media)
- other underlying medical disease (particularly cardiac and vascular disease); the more catabolic the underlying conditions, the greater the negative effect of surgery on renal function
- Age: elderly patients have a decreased number of functioning nephrons; however, their decreased muscle mass and creatinine production results in relatively lower glomerular filtration rates at any level of serum creatinine
- Emergency operations, bypass surgery, aortic (particularly aneurysmectomy) and renovascular surgery, and major intra-abdominal procedures (particularly in patients with gastrointestinal bleeding, gram-negative infections, and biliary obstruction); in general, serious intra-abdominal conditions have a greater negative effect on renal function than those above the diaphragm.

**Choice and Dosage of Drugs.** In patients with chronic renal disease, there must be a downward adjustment in the dosage of drugs that are metabolized by the kidney, those excreted unchanged, or those whose active breakdown products are cleared by this organ. Nephrotoxic drugs should be avoided, or if they are considered lifesaving, dosages, or frequency of administration, or both, must be reduced. The loading doses of most drugs remain little changed. The multiple formulas for reduced subsequent dosage schedules are directly related to the degree of reduction of the GFR and should be referred to for each drug. Although drug levels must be carefully followed, toxicity can occur with values in the therapeutic range. Reduced renal perfusion (volume depletion, hypotension) and any combination of such drugs potentiate the nephrotoxicity of all.

The aminoglycosides, methicillin, tetracycline, cephaloridine, amphotericin, and pentamidine are nephrotoxic. Other cephalosporins have little intrinsic nephrotoxicity, but may require reduced dosages in chronic renal failure and become increasingly toxic in combination with the aminoglycosides. Nafcillin is metabolized by the liver and can be used in these patients, but the synthetic penicillins may cause renal insufficiency on occasion. Many of the penicillins contain significant amounts of potassium or sodium, and the content of these electrolytes must be considered when they are used in these settings. Narcotics are degraded by the liver, but patients with renal failure are very sensitive to them, and when used, they should be given in small doses. Vancomycin may be given safely, but its dosage schedule must be adjusted for renal function. The potassium-sparing diuretics should not be used in the setting of chronic renal insufficiency. Ethacrynic acid, metolazone, and furosemide are safer and effective alone or in combination. The NSAIDs decrease glomerular filtration and can cause interstitial nephritis and should therefore be avoided. Depending on the underlying electrolyte abnormalities, care must be exercised in the use of antacids containing magnesium, aluminum, or calcium. The ACE inhibitors should be avoided because they can worsen renal function and potentiate hyperkalemia.

Radiocontrast dyes may cause a significant further decline in kidney function in patients with chronic renal disease and in those with dysproteinemias and diabetic nephropathy. If such a radiologic procedure is unavoidable, the patient must first be well hydrated.

The prophylactic efficacy of mannitol and furosemide prior to, during, and after the radiologic procedure, and of the calcium channel blockers nifedipine or verapamil given shortly before the procedure, remains to be proven.

## Perioperative Problems

Acute renal failure, hyperkalemia, hyponatremia, fluid overload, hypoventilation (prolonged effects of neuromuscular blocking agents), bleeding, pneumonia (gram-negative), and wound infections are common perioperative problems in patients with chronic renal insufficiency. In patients on hemodialysis, fistula thrombosis is frequently seen if care is not taken in positioning the access area and avoiding pressure on it. Cardiac arrhythmias resulting from electrolyte and pH disturbances are not infrequent.

Prerenal factors (hypotension and intravascular volume depletion) decrease renal blood flow and perfusion, potentiating the effects of all nephrotoxic insults.

**Prevention.** The key to management is prevention. Pivotal to the prevention of many complications (particularly acute renal failure) is preoperative hydration and intraoperative maintenance of the blood pressure and extracellular and intravascular volume with fluids as close to isotonicity as can be tolerated by the underlying renal or cardiac conditions. If blood is needed, fresh washed packed cells should be given. Dopamine may be used for poorly responsive hypotension. Many nephrologists recommend the prophylactic use of mannitol and furosemide in those patients undergoing high-risk vascular procedures (aortic and renovascular surgery) to protect the kidneys and perhaps prevent emboli.

**Stabilizing Central Venous Pressure.** Clinical assessments of volume status in many of these patients are often inaccurate. Whenever a serious concern exists, invasive monitoring with a central line or Swan-Ganz catheter should be undertaken and central venous pressure should be stabilized below 14–15 cm $H_2O$ or pulmonary capillary wedge pressures under 16–18 cm $H_2O$. Orthostatic hypotension and unexplained tachycardia in a patient with flat neck veins may be the earliest signs of volume depletion. Alterations in body weight, intake, and output must be carefully monitored. The textbook features—dry mucous membranes, altered skin turgor, soft eyeballs—are often late findings and may be misleading, particularly in the elderly. The presence of edema does not necessarily exclude volume depletion, because patients with congestive failure, nephrosis, and cirrhosis may have decreased effective circulating blood volumes in the presence of edema and are prone to hypotension from vigorous diuresis, blood loss, sepsis, and antihypertensive therapy. As a general rule, if the GFR is greater than 15 mL/min, edema is unlikely to be the result of renal failure itself, and other causes (hypoalbuminemia, etc.) should be sought. With full appreciation that volume-related hypertension is common in chronic renal disease and that these patients also have a high coincidence of underlying atherosclerotic heart disease and diminished left ventricular function (stiff ventricles), the fact remains that aggressive hydration prevents acute renal complications and saline should be given, unless contraindicated by left ventricular failure.

Much of the hypertension seen in patients with chronic renal disease is secondary to expanded extracellular volume and is responsive to loop diuretics or dialysis. Reasonable preoperative control should be attained as described above, or with other antihypertensives that do not decrease renal blood flow. The calcium channel and beta blockers are safe and usually effective, but the angiotensin-converting enzyme (ACE) inhibitors are best avoided. The high incidence of associated arteriosclerotic heart disease is not well understood and may be related to factors other than the hypertension, anemia, associated lipid abnormalities, and so forth seen in chronic renal disease. Careful preoperative evaluation of cardiovascular status is essential.

### Postoperative Acute Renal Failure

Up to 50% of patients with acute postoperative renal failure are nonoliguric. Frequent serum creatinine determinations will assist in early recognition.

Obstructive uropathy (prostate, anticholinergic drugs, retroperitoneal or pelvic tumors, bilateral nephrolithiasis, clogged catheter) must be excluded. Anuria, particularly in conjunction with intermittent polyuria, should suggest the possibility of obstruction; abdominal sonography can largely exclude it.

**Differentiating Prerenal Azotemia and Acute Tubular Necrosis.** Microscopic examination of a stained sediment of urine may reveal pyuria with eosinophilia, which suggests acute interstitial nephritis. Brick-red tubular casts or cells may be seen in acute tubular necrosis. A positive test for blood in the absence of red blood cells indicates that either myoglobin or hemoglobin may be present in the urine. As one evaluates the electrolytes, BUN, creatinine, osmolality, and the urinary sediment in the presence of chronic renal disease, the differentiation between prerenal azotemia and acute tubular necrosis becomes blurred. The characteristics of prerenal disease with a high urine osmolality (>500 mOsm), low urine sodium concentration (<20 mEq/L), and high ratio of serum BUN to creatinine (>20:1) may not be clear-cut in the presence of chronic renal disease. Frequently, the differential diagnoses cannot be made and a therapeutic trial of rehydration with normal saline must be carried out. Any question as to the preexisting level of hydration or of underlying heart failure may be clarified in clinically serious situations by invasive monitoring and evaluation of pulmonary capillary wedge pressure (PCWP). If the PCWP is low, the infusion should be continued, replacing the insensible losses as well as the hourly output of fluid and electrolytes. If rehydration has been attained without fluid overload (normal PCWP) and renal function does not improve, it may be presumed that further renal damage has occurred. The intravenous infusion may be discontinued and furosemide (single, large dose of 100–400 mg) should be given alone or in combination with zaroxolyn in an attempt to induce diuresis. Mannitol administration can cause increased plasma osmolality with a shift of water to the extracellular and intravascular spaces, resulting in dilutional hyponatremia and vascular congestion that could overload a compromised heart. The maintenance of diuretic-induced diuresis does not appear to influence the outcome of acute renal failure, but does facilitate management of the attendant fluid and electrolyte abnormalities.

The indications for dialysis are the same in oliguric as in nonoliguric patients. The use of hypertonic glucose and essential L-amino acids has been advocated by some nephrologists. Such therapy does appear to improve the BUN, phosphorus, calcium, and acidosis and may decrease dialysis requirements, but is associated with an increased incidence of sepsis, electrolyte abnormalities (hypokalemia and hypomagnesemia), and volume overload.

### Hyperkalemia

In patients with chronic renal disease and GFRs > 15 mL/min, preoperative hyperkalemia is most likely due to extrarenal factors, such as increased potassium intake, extracellular fluid volume depletion, rhabdomyolysis, absorption of hemolyzed blood, drugs (potassium-sparing diuretics, ACE inhibitors), or acidosis. Exceptions to this are patients susceptible to renal tubular disorders (those with, for example, the tubulointerstitial nephropathies of hyporeninemic hypoaldosteronemia, diabetes, obstruction, sickle cell disease). Such patients are vulnerable to hyperkalemia in the presence of only moderate reductions of GFR. As discussed above, multiple factors associated with surgery and anesthesia favor the development of hyperkalemia. These factors, in combination with the increased endogenous potassium production of the postoperative catabolic state, make symptomatic hyperkalemia in these patients a common occurrence.

Hyperkalemia is the most life-threatening perioperative electrolyte abnormality. Although electrocardiographic changes rarely occur with potassium levels below 6 mEq/L, all but the most lifesaving surgery should be deferred until serum potassium is reduced to less than 5 mEq/L. Kayexalate, by mouth or enema, removes potassium from the body and is the treatment of choice, if time permits. If this therapy is unsuccessful, dialysis should be undertaken. Glucose and insulin or bicarbonate (independent of its effects on pH) work rapidly, but only transiently by accelerating potassium transfer into the cells. These agents are useful in emergencies or as adjuvants to other therapy. Calcium, intravenously, almost immediately blocks the effects of potassium on the myocardial cell membrane and is useful in therapy of cardiac arrhythmias associated with hyperkalemia. Succinylcholine causes transient hyperkalemia (potassium is released from muscle during depolarization) and should not be used in this group of patients.

### Hypokalemia

Hypokalemia occasionally occurs as a result of massive losses via vomiting, diarrhea, and enteral tubal or fistula drainage. It is often associated with hypomagnesemia, especially in the setting of gentamicin administration, hyperalimentation, and malabsorption syndromes.

Replacement must be slow and cautious in these patients (less than 10 mEq/hr), via a peripheral vein, and with an understanding of the inverse relationship of pH and serum potassium. For example, as one corrects metabolic acidosis with bicarbonate, potassium moves intracellularly, which may accentuate the hypokalemia unless the electrolyte is given simultaneously.

Hypokalemia in the presence of alkalosis calls for use of the chloride salt for replacement, because acetate, citrate, and gluconate are metabolized to bicarbonate, further aggravating the alkalosis.

### Hyponatremia

Hyponatremia is usually not a preoperative problem in patients with GFRs > 15 mL/min. When hyponatremia occurs in such patients, correction should be based on its severity and its probable cause. Clinical problems rarely arise when the serum sodium concentration is greater than 125–130 mEq/L and are usually apparent only when concentration falls below 120 mEq/L.

If hyponatremia is associated with extracellular fluid volume (sodium) depletion (as from fluid and electrolyte losses from the kidney or gastrointestinal tract accompanied by hypotonic rather than isotonic fluid replacement), the patient should be treated with normal saline and restriction of hypotonic fluid intake.

If the patient is edematous (CHF or hypoalbuminemia secondary to liver disease or the nephrotic syndrome), the hyponatremia is best managed by correcting the underlying problem and applying intense fluid restriction.

If the etiology is SIADH, fluid restriction is the mainstay of therapy and the use of hypertonic saline, lasix, or declomycin as adjuvants depends on the acuity and severity of the condition.

Postoperative intensification or development of hyponatremia in patients with chronic renal disease and an impaired ability to handle free water is multifactorial, including (a) multiple surgical and anesthetic factors, particularly volume depletion, which favor the continued appropriate secretion of antidiuretic hormone, (b) the dilutional effects of intraoperative replacement of isotonic fluid losses with hypotonic solutions, and (c) increased water metabolism in the highly catabolic postoperative state.

If the patient is hypovolemic (volume status often requires invasive monitoring for certainty), management is replacement with normal saline; if euvolemic, hypotonic fluid restriction; and if hypervolemic, dialysis may be necessary.

### Other Electrolytes

It is not the purpose of this article to review the entire subject of electrolyte balance; the reader is referred to one of the many long review articles or texts on the subject.

**Hypocalcemia.** Hypocalcemia in chronic renal disease is usually asymptomatic because the associated acidosis increases the amount of ionized calcium. Correction of the acidosis may reverse this and can result in the development of symptomatic hypocalcemia. The aminoglycosides increase urinary magnesium, as well as calcium, losses. The administration of large amounts of citrated blood (5 or more units) may also aggravate an underlying asymptomatic hypocalcemia.

**Hypomagnesemia.** The clinician should suspect hypomagnesemia in any patient exhibiting neuromuscular or central nervous system hyperactivity postoperatively, particularly if the individual is an alcoholic. It should be surmised in the presence of refractory hypocalcemia, refractory hypokalemia (kaliuresis occurs with alkalosis or decreased serum magnesium), and intense aminoglycoside administration, and in patients who have had prolonged hyperalimentation.

### Acidosis

Acidosis causes myocardial depression, increases serum potassium, and predisposes to cardiac arrhythmias. Most patients with chronic renal disease are somewhat acidotic. As the GFR falls below 25 mL/min and the ability to form ammonia and other compensations break down, organic acids are retained and a high anion-gap acidosis ensues. Bicarbonate levels are usually maintained at 15–18 mEq/L and pH above 7.25 until the GFR falls below 10–15 mL/min. These are the minimum acceptable levels prior to surgery.

In the perioperative period, any metabolic insult that can cause a superimposed acidosis may result in a hyperacute fall in bicarbonate and pH levels. General anesthesia impairs respiratory compensations to acidosis and may cause a further fall in pH. Preoperatively the acidosis should be corrected with bicarbonate to attain a pH of at least 7.25. If limitations on electrolyte replacement exist, or if there is fluid overload, the patient should be dialyzed.

### Anemia

Patients with long-standing anemias having hematocrit levels of 25% generally tolerate surgery quite well without transfusions. The hazards of volume overload or hyperkalemia usually outweigh any slight benefit such transfusions might offer.

More severe anemias should be corrected to these levels with washed, fresh, leukocyte-poor, packed red cells.

If the contemplated surgery is associated with large blood loss, if there is underlying atherosclerotic heart disease or severe chronic pulmonary or peripheral vascular disease, or if it is difficult to correct the coagulopathy, prudence dictates bringing the hematocrit up to 30% prior to operation.

### Coagulopathy

The bleeding tendency in uremia is poorly understood. There are deficiencies in platelet aggregation and function, perhaps in the production or function of Von Willebrand factor, and possibly in vessel wall prostaglandins. Platelet counts, prothrombin time, and partial prothrombin time are usually normal, but the bleeding time becomes elevated. Such elevations are unusual until the GFR falls below 15–20 mL/min, BUN exceeds 90 mg/dL and creatinine exceeds 6 mg/dL. If the bleeding time exceeds 1.5 times normal, preoperative correction should be accomplished by dialysis. If correction is not complete with this modality, cryoprecipitate, which has its peak effect at 4–8 hr, or desmopressin acetate (DDAVP), which works in an hour (with effects lasting 6–8 hr) may be utilized. Conjugated estrogens have more prolonged effects on the bleeding time via unknown mechanisms, but the long latency before these agents take effect limits their usefulness in acute situations.

### Suggested Reading

Cockcroft DW, Gault MH. Prediction of creatinine clearance from serum creatinine. Nephron 1969; 16:31

Rowe JW. The influence of age on renal function. Resid Staff Physician 1978; 24(Feb):49–55.

Burke JF, Francos GC. Surgery in the patient with acute or chronic renal failure. Med Clin North Am 1987; 79:489.

Blythe WB. The management of intercurrent medical and surgical problems in the patient with CRF. In: Early, Gottschalk, eds. Strauss and Welt's diseases of the kidney. Boston: Little, Brown, 1979.

Brenowitz JB, Williams CD, Edwards WS. Major surgery in patients with chronic renal failure. Am J Surg 1977; 134:765.

Hampers CL, et al. Major surgery in patients on maintenance hemodialysis. Am J Surg 1968; 115:747.

Tilney NL, Lazarus JM. Acute renal failure in surgical patients. Surg Clin North Am 1983; 63:357.

Anderson RJ, Schrier RW. Clinical spectrum of oliguric and nonoliguric acute renal failure. In: Brenner, Stein, eds. Contemporary isues in nephrology: acute renal failure. New York: Churchill Livingstone, 1980.

Brown CB, et al. High dose furosemide in acute renal failure: a controlled trial. Clin Nephrol 1981; 15:90.

Kleinknecht D, Graneval D. Preventative hemodialysis in the treatment of acute renal failure. Its effect on mortality and morbidity. In: Friedman EA, Eliahou HE, eds. Proceedings of Conference on Acute Renal Failure. New York: DHEW Publications No. (NIH) 1973; 74-608:165.

Miller TR, et al. Urinary diagnostic indices in acute renal failure: a prospective study. Ann Intern Med 1978; 89:47.

Briefel GR, Turer P. Renal, fluid and electrolyte, and acid-base disorders. In: Kammerer W, Gross R, eds. Medical consultation. Baltimore/London: Williams & Wilkins, 1983.

Tasker PRW, et al. Prophylactic use of IV saline in patients with chronic renal disease undergoing major surgery. Lancet 1974; 2:911.

Janson PA, et al. Treatment of the bleeding tendency in uremia with cryoprecipitate. N Engl J Med 1980; 303:1318.

Mannucci PM, et al. dDAVP shortens the bleeding time in uremia. N Engl J Med 1983; 315:731–735.

Miller RD, et al. Succinylcholine-induced hyperkalemia in patients with renal failure. Anesthesiology 1972; 36:138.

Weir PHC, Chung FF. Anesthesia in patients with chronic renal disease. Can Anaesth Soc J 1984; 31:468.

Fernandez P, Cox M. Basic concepts of renal physiology. The kidney in anesthesia. Int Anesthesiol Clin 1984; 22:1–34.

Deutsch S. Anesthetic effects on the kidney. Kidney 1971; 4:1.

Bennett WM, et al. Drug prescribing in renal failure: dosing guidelines for adults. Am J Kidney Dis 1983; 3:155.

Harkonen S, Kjellstrand CM. Contrast nephropathy. Am J Nephrol 1981; 1:69.

Clive DM, Stoff JS. Renal syndromes associated with NSAID's. N Engl J Med 1984; 310:563.

Francis GS, et al. Coronary artery disease in patients with end stage renal disease. Ann Intern Med 1980; 92:499.

Ayus JC, Fromme JP, Young JB. Cardiac and circulatory abnormalities in chronic renal failure. Semin Nephrol 1981; 1:112.

Brenner BH, Rector FC Jr. The kidney, 4th ed. Philadelphia: Saunders, 1990.

Seldin DW, Grebisch G. The kidney. New York: Raven Press, 1985.

Brenner BM, Lazarus JM. Acute renal failure, 2nd ed. Philadelphia: Saunders, 1990.

Kokko JP, Tannen RL. Fluids and electrolytes, 2nd ed. Philadelphia: Saunders, 1990.

Arieff AI. Hyponatremia. Mt Sinai J Med 1990; 57:125–135.

Sterns RH. The treatment of hyponatremia: unsafe at any speed? AKF Nephrol Lett 1989; 6(1):1–10.

# Preoperative Evaluation of the Patient with Liver Disease

Mitchell Conn, M.D.

Medical consultants are not uncommonly called on to evaluate a patient awaiting surgery who has unexpected or known liver disease. An estimated 1:700 patients admitted for elective surgery are found to have abnormal results on liver tests (1). The great concern is that the underlying liver disease will predispose to greater surgical risk and deterioration of liver function, leading to postoperative liver failure, complications, or death. The unique metabolic and synthetic role the liver plays, coupled with the effects of anesthesia and surgery on liver function, does predispose the patient with underlying liver disease to increased perioperative morbidity and mortality. A greater understanding of liver function and the effects of anesthesia on it is necessary to optimally assess and manage these patients. The medical consultant's advice must be based on an understanding of the pathophysiology and etiology of the liver condition, the functional capacity of the liver, and the urgency of the proposed surgery.

## Liver Anatomy and Function in Normal and Diseased States

The liver derives its blood supply from dual sources: 75% from the oxygen-poor, nutrient-rich branches of the portal vein and 25% from the hepatic artery.

The liver plays a vital role as a blood reservoir; it can deliver an additional 500 mL of blood into the circulation in the event of hemorrhage. Regulation of hepatic portal flow is predominantly under alpha-adrenergic control. Patients with underlying liver disease have a decreased sensitivity to catecholamines and consequently may have blunted responses to intraoperative hypovolemia or hemorrhage. Anesthetics that suppress the sympathetic nervous system may further inhibit this response.

The liver is divided into hexagonal lobules. At the angles are the portal triads and in the center is the terminal hepatic venule (central vein). The hepatocytes form plates radiating from the central vein to the portal triads. Between these plates of hepatocytes lie the sinusoids (vascular channels). Bile canaliculae are located between hepatocytes. The hepatocytes are divided into zones: zone 1 cells are closest to the portal triads and, being most proximal to the blood supply, receive oxygen first, are most resistant to necrosis due to hypoxia or anoxia, and regenerate earliest; zone 3 cells are most distal to the blood supply and are most susceptible to hypoxia and hepatotoxins.

The hepatocytes perform a multitude of functions. They store and secrete nutrients and bile salts; synthesize plasma proteins, glucose, and cholesterol; and metabolize and detoxify drugs, toxins, and multiple endogenous compounds. The major metabolic functions of the liver are the excretion of bilirubin, maintenance of normal blood glucose levels (store glycogen and release glucose), and deamination of amino acids with the formation of urea and removal of ammonia. Other functions include the production of multiple blood coagulation factors and the storage of iron and vitamins. The reticuloendothelial Kupfer cells phagocytize bacteria and other foreign matter as it passes through the liver.

These functions may deteriorate during surgery and anesthesia in both normal and hepatic-

---

From the Divisions of Gastroenterology and Hepatology, Department of Medicine, Mount Sinai School of Medicine, New York, NY. Address reprint requests to the author at Box 1069, Mount Sinai Medical Center, One Gustave L. Levy Place, New York, NY 10029.

ally impaired patients. The ability of the liver as a detoxifier may be affected by multiple mechanisms, chiefly by alterations in hepatic function and blood flow. Depletion of specific plasma proteins (for instance, albumin) results in altered binding and metabolism of many drugs. Patients with cirrhosis are highly sensitive to morphine, chlorpromazine, and diazepam. The central nervous system uptake of benzodiazepines is increased in hepatic encephalopathy owing to increased receptors. Elevated levels of glucagon in patients with cirrhosis suppress catecholamine responsiveness. Proper titration of such drugs is important in cirrhotic patients (2–6).

### Pathophysiology of Liver Disease

Liver disease can be divided into two broad categories, hepatocellular (parenchymal) and cholestatic.

Advanced cirrhosis affects the function of many essential organs and systems that may adversely affect surgical outcome.

- Cardiovascular function in advanced liver disease is characterized by a hyperdynamic state (high output, low peripheral resistance, and arteriovenous shunting) and reduced responsiveness to catecholamines. These patients are in a sodium-avid state, which leads to the accumulation of ascites and edema. The ascites may further compromise cardiovascular function by increasing intra-abdominal pressure, thereby reducing venous return to the heart. Chronic ethanol abuse may lead to alcoholic cardiomyopathy (a low-output congestive heart failure).
- Renal blood flow is well preserved until advanced liver disease develops. Electrolyte abnormalities are generally secondary to hyperaldosteronism and diuretic therapy. Reduction in renal cortical blood flow despite the high-output state is the first sign of renal dysfunction and the harbinger of the hepatorenal syndrome.
- Total hepatic blood flow is reduced, predominantly owing to decreased portal blood flow. This has paramount significance with regard to the metabolism of endogenous and exogenous compounds and drugs. Prior portacaval shunt surgery further reduces portal flow to the liver.
- There is an increased content of red blood cell 2,3-DPG with a resultant shift in the oxygen dissociation curve to the right (decreased affinity of hemoglobin for oxygen).
- Hypoxia may develop due to intrapulmonary shunts and to hypoventilation secondary to ascites and decreased diffusing capacity (expanded extracellular volume).
- There may be multiple associated blood and coagulation abnormalities. Anemia may be secondary to increased plasma volume, gastrointestinal blood loss (varices, gastritis, ulcer), nutritional deficiencies, and hemolysis (splenomegaly, chronic disease). Leukopenia and thrombocytopenia may also be observed due to the hypersplenism. Reduction in the concentration of hepatically synthesized coagulation factors leads to prolongation of the prothrombin and partial thromboplastin times.
- The decreased volume of hepatic cells and their inability to synthesize and metabolize glycogen, insulin resistance, and alterations in glucagon and growth hormone metabolism eventuate in glucose intolerance in these patients.

Cholestatic liver disease may be either extra- or intra-hepatic, due to bile duct obstruction or to primary hepatic pathology. It results in the release into the circulation of substances usually excreted into the bile (bilirubin, bile acids, cholesterol, and so on). Cardiovascular changes are similar to, but much less than, those seen in hepatocellular disease. Coagulation abnormalities are usually secondary to deficiency of vitamin K–dependent coagulation factors and are correctable by administration of the vitamin. Only after long-standing cholestasis does sufficient hepatocellular dysfunction result so as to decrease other coagulation factor synthesis (4–6).

### Effects of Anesthesia on the Liver

The various anesthetics affect hepatic function differently, but most of the inhalation agents decrease hepatic blood flow. Postoperative hepatic dysfunction may be due to this decreased flow or to the direct hepatotoxic effects of these drugs.

Halothane has well-recognized hepatotoxicity. Fatal hepatic necrosis is rare (1:35,000), but severe hepatic dysfunction may occur in 1:6,000. Risk factors include multiple exposures to the drug, obesity, female gender, and middle age. The toxic mechanism of halothane is not known but is probably immunologic. Beyond its ability to cause hepatitis, halothane affects liver function to a greater degree than any of the inhalation agents. It alters synthetic and excretory function and decreases hepatic oxygen supply and total hepatic blood flow. If a question exists with regard to liver function, use of halothane should be avoided.

Enflurane and isoflurane have only rarely been associated with postoperative liver dysfunction; a causal relation is, at most, only suspected. The effect of these two agents on hepatic blood flow and oxygenation is much less than that of halothane. They are safe to use in patients who have liver disease. Nitrous oxide has no hepatotoxicity (6).

Intravenous anesthetics and opiates probably do not affect hepatic function. However, clearance of commonly used agents such as midazolam, fentanyl, and morphine may be delayed in cirrhotics, and dosing may require adjustment.

The pharmacodynamics of several drugs is altered in liver disease. The half-life of lidocaine is increased 300%, of the benzodiazepines, 100%. Drugs bound to albumin (pentothal) should be given in reduced dosages in hypoalbuminemic states. The muscle relaxants atraciurium and vecuronium may be used in liver disease (4, 6).

Patients with well-compensated liver disease tolerate most drugs well. When dealing with decompensated patients, drug dosage must be adjusted. A good rule of thumb is to reduce the dose by 50% initially and then adjust as tolerated and needed.

## Effects of Surgery on Hepatic Function

The stresses of surgery may induce changes in liver hemodynamics leading to hepatic dysfunction. Exploratory laparotomy results in reduced blood flow to the intestines and liver, whereas minor surgical procedures are unlikely to cause such changes. This effect is most likely related to stress-induced release of catecholamines and vasopressin and activation of the renin-angiotensin system. In patients with advanced liver disease, alterations in hepatic blood flow can result in worsening hepatic function and increased morbidity. Anesthetics and surgical intervention result in an additive reduction in hepatic blood flow; the magnitude of the impact depends on the nature of the surgical intervention and the type of anesthetic. Upper abdominal surgery is associated with the greatest reduction in hepatic blood flow; best-choice inhalation anesthesia is isoflurane, which has the least effect on hepatic blood flow and oxygenation (6, 7).

Elevations in liver function tests are more likely to occur after biliary tract surgery than after nonabdominal procedures. Intraoperative factors such as hypotension, hypoxia, hemorrhage, sepsis, and congestive failure may also affect hepatic blood flow (7).

## Surgical Risk in Patients with Liver Disease

Surgical risk in patients who have liver disease depends largely on the type and severity of the liver disease and the extent of the surgery.

Patients with acute viral hepatitis are at increased surgical risk. Harville and Summerskill reported 10% mortality and 12% significant morbidity in patients with viral hepatitis undergoing laparotomy. This correlated with the extent of surgery (8). Patients with drug-induced hepatitis did well at surgery. Turner and Sherlock reported an incidence of 42% mortality and 33% serious hepatic decompensation in such patients undergoing surgery (9). These early observations suggest that in patients with acute viral hepatitis, elective surgery should be postponed until the patient recovers clinically and chemically. Clinically, one should suspect an acute viral hepatitis when a patient initially has moderate to marked elevations of transaminase (more than 3 times normal, usually much higher) associated with fever, nausea, malaise, right upper quadrant pain, and jaundice. In patients with subclinical disease, the diagnosis may not be as evident.

Chronic hepatitis is defined as elevation in liver enzymes persistent for six months. Pathologically, this may arise from a number of disorders; liver biopsies done on such patients reveal several origins, including alcoholic liver disease, fatty liver, chronic viral hepatitis, and miscellaneous (drug-induced, autoimmune, primary biliary cirrhosis) (10). There are no studies of the risk of surgery in patients with chronic hepatitis. These patients should be evaluated on the basis of liver function. If they are asymptomatic with good synthetic capacity, there should be no significant increased risk. If they are symptomatic with poor synthetic function, all but emergent surgery should be avoided until the patient is clinically improved.

Patients with alcoholic liver disease represent a spectrum ranging from fatty liver, to acute alcoholic hepatitis, and finally to cirrhosis. Alcoholic fatty liver is generally subclinical and not associated with significant liver dysfunction. Surgery can be performed on these patients safely. Acute alcoholic hepatitis is characterized by significant inflammation and necrosis of liver cells. Patients have an initial clinical appearance similar to patients with viral hepatitis. However, the transaminases are generally not as high

TABLE 1
*Risk Factors for Surgery in Patients with Cirrhosis*

| Factor | Mortality with (%) | Mortality without (%) |
|---|---|---|
| Ascites | 37–83 | 11–53 |
| Albumin <3 mg/dL | 58 | 12 |
| Bilirubin >3 mg/dL | 44–62 | 13–17 |
| Prothrombin time >1.5 sec control | 63 | 18 |
| Emergency surgery | 45–86 | 10–40 |
| Abdominal surgery | 35 | 10 |
| Preoperative infection | 64 | 21 |
| Child's class | | |
| A | 5–10 | 50 |
| B | 31 | 30 |
| C | 76 | 18 |

Adapted from Refs. 11–13.

(<500) and the SGOT is more elevated than the SGPT. The disease carries a high mortality. Elective surgery should be avoided and should be considered only after a prolonged period of abstinence (2–3 months), clinical improvement, and normalization of the liver enzymes. Patients with alcoholic cirrhosis should be evaluated and managed like other patients with cirrhosis.

Surgery in cirrhotics is associated with greater postoperative mortality. The most common cause of death postoperatively are liver failure, sepsis, bleeding, and renal failure. The key factors in predicting outcome in such patients are the level of hepatic function and the type and urgency of the surgery. Preexisting ascites, hypoalbuminemia, hyperbilirubinemia, or coagulopathy, abdominal and/or emergency surgery, preoperative infection, and Child's class B and C (Table 1) (11–13) are all associated with poor outcomes. One study (14) demonstrated an excellent correlation between the aminopyrine breath test and surgical outcome; 97% of patients with a normal test survived, whereas 86% of patients with abnormal test results died.

### Preoperative Evaluation and Management

During the preoperative evaluation of a patient with abnormal liver enzymes, the medical consultant should ascertain any history of recent drug use, past history of hepatitis or jaundice, history of blood transfusions, and any clinical symptoms of liver disease (malaise, fatigue, arthralgias). Physical examination should focus on stigmata of chronic liver disease (cutaneous spider angiomata, parotid swelling, gynecomastia, palmar erythema, Dupuytren's contractures, organomegaly, ascites, encephalopathy, poor nutritional status). The presence of xanthelasma should suggest primary biliary cirrhosis.

Blood tests should include a complete blood count with platelets, blood urea nitrogen (BUN) and creatinine, electrolytes, total protein, albumin, globulin, bilirubin, alkaline phosphatase, SGOT, SGPT, LDH, prothrombin time, and, when indicated, serum ammonia levels. The SGOT and SGPT in chronic liver disease are considerably lower than in acute hepatitis and may be normal; they reflect the extent of acute liver injury rather than hepatic function in this circumstance. The SGPT is felt to be more specific to the liver than the SGOT. Elevation of the alkaline phosphatase is reflective of cholestatic disease, not hepatocellular injury or altered synthetic function. Since this enzyme is found in bone, in placenta, and in the intestines, even more specific tests for cholestasis are the gamma-glutamyl transpeptidase and 5′-nucleotidase.

Preoperative assessment must determine the functional ability of the liver. Measurement of serum proteins synthesized by the liver (for example, albumin) or of the prothrombin time, which indicates a decrease in hepatic synthesis of coagulation factors (in the absence of vitamin K deficiency, ingestion of Coumadin, disseminated intravascular coagulation, etc.), is a far more sensitive indicator of hepatic function. Low BUN levels suggest a decreased ability to synthesize urea. Hyperglobulinemia is usually present in advanced liver disease. Thrombocytopenia may reflect either hypersplenism from portal hypertension or alcohol-induced bone marrow suppression. Surgery in patients who rank as Child's class B or C (Table 2) should be delayed or avoided if possible, at least until correction and improvement in liver function and class occurs. If the patient feels and looks well, a mild, stable abnormality in transaminases alone (less than 3 times normal) should not prohibit surgery. If there is doubt as to the type and extent of liver disease, a liver biopsy should be performed.

Ischemic hepatitis is a specific form of hepatitis of which the consultant should be aware. It

TABLE 2
*Child's Classfication*

| Factor | Class A | Class B | Class C |
|---|---|---|---|
| Bilirubin (mg/dL) | <2.0 | 2.0–3.0 | >3.0 |
| Albumin (mg/dL) | >3.5 | 3.0–3.5 | <3.0 |
| Ascites | None | Controlled | Refractory |
| Encephalopathy | None | Mild | Advanced |
| Nutrition | Excellent | Good | Poor |

Adapted from Ref. 15.

usually occurs along with hypotension in patients with cardiac disorders and very ill patients. Centrilobular necrosis is seen histologically. The condition mimics acute hepatitis, with significant elevations of transaminases but only mild rises in bilirubin and alkaline phosphatase. High levels of LDH are characteristic. Severe hepatic dysfunction generally does not result, and the condition usually resolves once the patient has been returned to hemodynamic stability.

Preoperative treatment for patients with severe liver disease should be devoted to correction of coagulopathy and electrolyte abnormalities and control of ascites, hepatic encephalopathy, and infection. The prothrombin time should be returned to within 3 sec of the control—if time permits, by using parenteral vitamin K (10 mg subcutaneous or intramuscular); or by the use of fresh frozen plasma. Lack of correction of the prothrombin time with vitamin K confirms hepatic parenchymal dysfunction and bodes a poorer prognosis. Some patients respond to desmopressin acetate (DDAVP). The platelet count should be at least 100,000. Electrolyte abnormalities, if severe, should be corrected. Hypokalemia and hyponatremia (usually from diuretic use) are the most commonly found in cirrhotics. If massive ascites is present, attempts should be made to reduce it. Paracentesis may be necessary to exclude infection or malignancy. Salt and free water intake should be restricted, and intravenous fluid replacement should include colloid, such as salt-poor albumin or fresh frozen plasma. Patients with end-stage liver disease are susceptible to postoperative renal failure. Intravascular volume should be optimized (with colloid), and nephrotoxic agents (aminoglycosides, NSAIDs) should be avoided. Encephalopathy should be corrected preoperatively. Conventional therapy includes correction of the precipitating causes (gastrointestinal bleeding, electrolyte imbalance, infection, constipation, and drugs), lactulose or neomycin (if renal function is unimpaired), and reduction of total and animal protein in the diet.

## Postoperative Management

The guidelines for postoperative patient care are analogous to those for preoperative management. Close monitoring and correction of coagulopathy, electrolyte abnormalities, fluid balance, hepatic encephalopathy and infection are essential.

Postoperative liver dysfunction in patients without preoperative liver disease can result from acute viral hepatitis, exacerbation of unsuspected chronic hepatitis, anesthesia-induced hepatitis (halothane), drug-induced hepatitis, infection, blood transfusions, intraoperative hepatic hypoxia, and hypotension. Postoperative jaundice may be secondary to increased bilirubin load (secondary to blood transfusions, hemolysis, hematomas), intrahepatic cholestasis secondary to hepatocellular injury or drugs, and extrahepatic biliary obstruction (gallstones, strictures, pancreatitis). Most often such jaundice is self-limited and will resolve spontaneously. Preexisting liver disease probably plays a substantial role in postoperative liver dysfunction (7).

## Summary

Patients with underlying liver disease have an increased surgical risk. Prediction of surgical risk is based on the degree of liver dysfunction, the type of surgery, and the preoperative clinical status of the patient. The medical consultant must determine the type of liver disease (acute or chronic) and the degree of impairment of liver function (based on physical examination and laboratory data). Any abnormalities should be corrected prior to surgery. Elective surgery should be postponed in patients with decompensated cirrhosis (class B and C) and acute hepatitis until the patient improves. The consultant internist's care of the patient should continue during the early postoperative course to monitor any deterioration in liver function.

### References

1. Schemel WH. Unexpected hepatic dysfunction found in multiple laboratory screening. Anesth Analg 1976; 55:810.
2. Jones AL. Anatomy of the normal liver. In: Zakim D, Boyer TD, eds. Hepatology. A textbook of liver disease. Philadelphia: Saunders, 1990.
3. Withrington PG, Richardson PDI. Liver blood flow. In: Zakim D, Boyer TD, eds. Hepatology. A textbook of liver disease. Philadelphia: Saunders, 1990.
4. Bass NM, Williams R. Hepatic function and pharmacokinetics. In: Zakim D, Boyer TD, eds. Hepatology. A textbook of liver disease. Philadelphia: Saunders, 1990.
5. Jenkins WJ, Billing B. Physiology of the liver. In: Berk JE, ed. Bockus gastroenterology, liver, Vol 5, Philadelphia: Saunders, 1985.
6. Gelman S. Anesthesia and the liver. In: Barash PG, Cullen BF, Stoetting RK, eds. Clinical Anesthesia. Philadelphia: Lippincott, 1989.
7. Farrell GC. Postoperative hepatic dysfunction. In: Zakim D, Boyer TD, eds. Hepatology. A textbook of liver disease. Philadelphia: Saunders, 1990.
8. Harville DD, Summerskill WHJ. Surgery in acute hepatitis. Cause and effects. JAMA 1963; 184:257.
9. Turner MD, Sherlock S. In: Smith R, Sherlock S, eds. Surgery of the gallbladder and bile ducts. London: Butterworth, 1964.

10. Van Ness MM, Diehl AM. Is liver biopsy useful in the evaluation of patients with chronically elevated liver enzymes? Ann Intern Med 1989; 111:473.
11. Garrison RN, Cryer HM, Howard DA, Polk HC. Clarification of risk factors for abdominal operations in patients with hepatic cirrhosis. Ann Surg 1984; 199:648.
12. Doberneck RC, Sterling WA Jr, Allison DC. Morbidity and mortality after operation in nonbleeding cirrhotic patients. Am J Surg 1983; 146:306.
13. Aranha GV, Greenlee HB. Intra-abdominal surgery in patients with advanced cirrhosis. Arch Surg 1986; 121:275.
14. Gill RA, Goodman MW, Golfus GR, Onstad GR, Bubrick MP. Aminopyrine breath test predicts surgical risk for patients with liver disease. Ann Surg 1983; 198:701.
15. Child CG, Turcotte JG. Surgery and portal hypertension. In: Child CG, ed. The liver and portal hypertension. Philadelphia: Saunders, 1964.

# The Pregnant Patient: Need for Surgery Unrelated to Pregnancy

Sheldon H. Cherry, M.D.

Although the frequency of nonobstetric surgery on pregnant patients is low, when the need for such procedures arises certain basic principles are important. The disease requiring surgery and the amount of pathology—not the surgery itself—are the most important factors in determining the effect on maternal and fetal wellbeing and morbidity. Perinatal mortality does not significantly increase for well-planned procedures carried out in the second trimester. However, fetal and maternal morbidity and mortality increase in incidence with increasing severity of the underlying surgical problem, especially if that problem is associated with sepsis, shock, hypotension, or hypoxia. One problem is the confusion in diagnosis due to the altered physiologic status of the woman who is pregnant. In addition, anatomic alterations play a part in delaying intervention.

## Physiologic Factors in Pregnancy

The physiology of pregnancy differs from that of the nonpregnant state. Many of these adaptive changes are important for maintaining fetal and maternal needs. An increase in red blood cell mass and intravascular volume (40%) provides for the increased profusion requirements of the fetal-maternal interface and also protects the mother in case of blood loss as a result of trauma or at delivery. The maternal vascular space accommodates this increase in blood volume without the development of hypertension. The increase in blood volume causes increasing renal blood flow and glomerular filtration rate.

Cardiovascular changes include increased oxygen consumption, increased heart rate, and increasing cardiac output, with a peak effect at 28 weeks of gestation. Blood pressure falls during the second trimester, with a greater decrease in the diastolic pressure resulting in an increase in pulse pressure. The increase in blood volume occurs to a greater extent in the plasma component than in the red blood cell mass, resulting in physiologic anemia even though the red blood cell mass is somewhat increased.

Other physiologic adjustments that occur during pregnancy are altered glucose homeostasis and maternal respiratory alkalosis. The compensated maternal respiratory alkalosis is probably due to hyperventilation, which begins early in pregnancy and is thought to be due to increased progesterone as well as sensitivity of the central nervous system.

As the gravid uterus elevates the diaphragm, the functional reserve capacity of the lung falls. Pregnant women also complain of dyspnea with a reduction of exercise tolerance. In addition to the increase in total blood volume and the related physiologic anemia, there is a leukocytosis normally of up to 15,000. This physiologic factor is taken into account in evaluating an acute surgical patient. The Table summarizes some of the physiologic alterations in pregnancy and their relevance to the surgical patient.

## General Concepts

Pregnancy is unique because the physician is dealing with two patients: the woman and the fetus. Disease, treatment, and pathology affect both. Sometimes the potential benefits to one may produce some hazards to the other. One ex-

---

From the Department of Obstetrics and Gynecology, the Mount Sinai School of Medicine, New York, NY 10029. Address reprint requests to the author at Box 1170, Mount Sinai Medical Center, One Gustave L. Levy Place, New York, NY 10029.

TABLE

*Physiologic Alterations in Pregnancy and Their Relevance to the Surgical Patient*

| System change | Clinical implications |
|---|---|
| **Cardiovascular** | |
| Uterine compression of vena cava and aorta in supine position | Decreased cardiac output and uterine perfusion; tilt hip 15° in perioperative period |
| Decrease in blood pressure in early to mid-gestation | Altered criteria for diagnosis of hypotension |
| Presence of dyspnea, third heart sound, and edema | None. Normal feelings |
| **Respiratory** | |
| Decreased arterial $PO_2$ when patient is in supine position | Avoid supine recumbency |
| Decreased pulmonary functional residual capacity and increased $O_2$ consumption | Increased risk of hypoxia perioperatively. Avoid hypoventilation. Increase inspired $O_2$ content prior to procedures inducing apnea (intubation or tracheal suctioning) |
| **Hematologic** | |
| Decreased venous flow in legs and increased levels of clotting factors | Increased risk of thromboembolism; avoid supine position. Use support stockings or prophylactic heparin |
| Proximity of fetal and maternal circulations | Risk of isoimmunization; $Rh_o$ (D) immune globulin with uterine trauma |
| Physiologic anemia, leukocytosis | Altered norms |
| **Gastrointestinal** | |
| Decreased gastric motility and reduced competency of gastroesophageal sphincter | Increased risk of aspiration; preoperative antacids should be considered. Less stress ulcers |
| **Renal** | |
| Dilatation of urinary collecting system | Increased risk of urinary infection. Catheterization should be avoided |
| 30%–50% increase in glomerular filtration rate and renal plasma flow; concomitant decrease in serum creatinine and urea nitrogen | Serum creatinine above 0.8 mg/dL may reflect impaired renal function; clearance of many drugs increased |

ample might be the use of chemotherapy to treat a specific maternal neoplasm; the fetus may be harmed by the effects of the drug. Sometimes to achieve a better fetal outcome the clinician will treat a pregnant woman who unpregnant might not require therapy. An example of this would be gestational diabetes or idiopathic thrombocytopenic purpura. Pregnant women with severe sickle-cell disease might require exchange transfusions to maintain adequate placental profusion. However, in the nonpregnant state these transfusions would be avoided. It is a truism, however, that any acute process that affects the woman may ultimately harm the fetus. An example would be maternal trauma with a bone fracture, which ultimately could affect fetal well-being if sepsis or embolus occurs. Whichever of the two is the intended beneficiary of the treatment, the other may be affected. The goal of the clinician is to minimize complications and maximize therapy for both the woman and the fetus.

## Anesthesia in the Pregnant Patient

Routine anesthetic management of the pregnant patient also involves two patients, the woman and the fetus. Maternal factors affecting anesthesia include increased cardiac output, oxygen consumption, and glomerular filtration rate, as well as decreased gastric emptying, peripheral vascular resistance, and impaired venous return resulting from compression of the vena cava by the gravid uterus in the supine position.

The major risks of anesthesia to the fetus are maternal hypoxia, hypotension, and acidosis. However, little is known about the specific effects of anesthetic gases on the fetus. There is little support for one anesthetic agent over another with regard to the risk of abortion, prematurity, or congenital malformations, if good oxygen concentration and maternal blood pressure are maintained. As a general principle, if possible anesthesia is best avoided during early organogenesis. With spinal and epidural anesthetics, hypotension with subsequent decrease in placental blood flow leading to fetal hypoxia should be avoided. Generally this is successful. There is no good objective evidence that the routine use of tocolysis to prevent perioperative premature labor is of value. However, tocolytic agents are available if premature labor occurs.

## Common Surgical Conditions

**Appendicitis: The Acute Abdomen.** Appendicitis is the most common cause of acute abdomen in pregnancy, occurring once in every 1,500 pregnancies. Owing to alterations in the normal anatomic landmarks of the appendix as pregnancy progresses, as well as other factors, the diagnosis of appendicitis in pregnancy is frequently delayed. Nausea, vomiting, and loss of appetite are common symptoms of both appendicitis and early pregnancy. Appendicitis occurs with equal frequency in every trimester. Because of the delay in diagnosis of acute appendicitis, the rate of perforation is significantly higher in the pregnant woman than in the nonpregnant woman. The reasons for the delay in diagnosis are

- the altered anatomic position of the appendix
- the lack of "walling off" of the appendix by the omentum secondary to increased levels of steroids that may suppress the inflammatory response
- misdiagnosis due to altered laboratory test values
- pregnancy symptoms that may mimic appendicitis symptoms.

The differential diagnosis of acute appendicitis includes diseases of gynecologic origin, such as salpingitis, twisted ovarian cyst, degeneration of fibroids, and ectopic pregnancy, as well as other abdominal problems, such as cholecystitis, Crohn's disease, and intestinal obstruction. Perhaps the most frequent disease requiring differentiation is pyelonephritis, also common during pregnancy.

The treatment of acute appendicitis is the same for the pregnant and the nonpregnant woman.

A delay in surgery is more serious during pregnancy. The risk of surgery to both woman and fetus is low. However, the consequence of a perforated appendix is an increase in morbidity and mortality.

The site of incision for appendectomy alters during pregnancy and is usually a vertical paramedian incision over the area of maximum tenderness.

With appropriate diagnosis and surgery a good outcome can generally be expected.

**Intestinal Obstruction.** Intestinal obstruction during pregnancy is relatively rare; however, the diagnosis can be difficult:

- Vomiting is a symptom of obstruction but may be confused with hyperemesis of the first trimester and may be attributed to the pregnancy itself.
- Distention of the abdomen is also more difficult to evaluate because of the size of the uterus and the normal distention.
- Pain caused by obstruction of the intestine may sometimes be confused with uterine pain or labor.

The causes of intestinal obstruction in pregnancy, as in the nonpregnant state, are adhesions from previous surgery and volvulus. The presence of a surgical scar on the abdomen of a pregnant women with abdominal pain must alert the clinician to the possibility of obstruction.

The treatment of volvulus is the same in the pregnant patient as in the nonpregnant patient. Surgery is usually necessary unless a sigmoid volvulus can be decompressed via a sigmoidoscope.

Incarcerated hernias are rare causes of obstruction during pregnancy.

The question of abdominal x-ray and an obstructive x-ray series versus fetal x-ray exposure may arise. It is generally felt that x-rays are mandatory, despite the pregnancy, if the diagnosis of obstruction is seriously entertained.

The treatment of obstruction is the same as in the nonpregnant state: adequate fluid and electrolyte replacement. If the etiology is adhesions, the use of a long tube may be tried. Surgery is indicated if there is suspicion of acute changes.

**Gallbladder Disease.** Biliary tract disease during pregnancy is not rare. However, the need for surgery during pregnancy is quite uncommon. Gallbladder symptoms are often present in more than 50% of patients prior to pregnancy. Pregnancy predisposes to the formation of gallstones because of an increase in bile stasis and an increase in concentration of cholesterol in the bile. General smooth-muscle relaxation also leads to stasis and gallbladder dilatation. Gallbladder disease can occur in any trimester, but its incidence increases progressively during pregnancy. It is more common in multiparas than in primiparas.

The signs and symptoms of gallbladder disease are the same in the pregnant and nonpregnant state, with pain in the right upper quadrant radiating back to the ribs. Nausea and vomiting as well as tenderness in this region are common. Jaundice, if present, is particularly significant. Alkaline phosphatase may be normally elevated in pregnancy, but the transaminases are not.

The best test for the diagnosis of gallstones in pregnancy is clearly a sonogram, which is safe and highly accurate.

Management of the pregnant woman with cholecystitis rarely requires surgery and is generally conservative. The indications for surgery are lack of response to medical treatment or unrelenting pain, requiring a cholecystectomy. The optimal time for surgery, if necessary, is in the second trimester, when the size of the uterus is not a major technical problem and the possibility of abortion is less. However, the need for surgery is unusual. If pancreatitis complicates cholelithiasis, maternal and fetal morbidity increases.

**Peptic Ulcer Disease.** Peptic ulcer disease rarely becomes a problem during pregnancy. Most women with this disease improve during pregnancy because of decreased secretion of gastric acids as well as decreased gastrointestinal motility. However, occasionally a perforated duodenal ulcer may occur, requiring surgery.

Endoscopy is the most important diagnostic procedure.

Early diagnosis and treatment will usually lead to good outcome.

**Ulcerative Colitis.** Ulcerative colitis is a disease occurring in women during the childbearing years. Generally patients in remission prior to the pregnancy do well during the pregnancy. The worst cases are those that occur de novo during the pregnancy. The need for surgery is unusual; generally the drugs of choice include corticosteroids, sulfasalazine, and some antibiotics. The use of 6-mercaptopurine should be avoided during pregnancy, if possible.

**Trauma.** The problem of trauma in pregnancy has become increasingly important. Traumatic injuries are the most common cause of death in women between the ages of 15 and 45, this time of greatest risk coinciding with the childbearing years. Automobile accidents and other traumatic injuries can occur in pregnancy. Injuries specific to the uterus, fetus, and placenta can be particularly hazardous, and treatment must be individualized.

During acute treatment it is crucial to maintain adequate blood volume and oxygenation. Shock or hypotension must be treated promptly. The supine hypotensive syndrome due to pressure on the vena cava must be kept in mind because it can lead to a loss of up to 30% of the effective circulating blood volume.

The most important principle in treating the injured pregnant woman is to maintain her cardiac output and maintain her blood or plasma volume.

## Summary

A team approach combining the skills of obstetrician, internist, and surgeon is required to optimize care for both woman and fetus if an acute surgical event should occur. A well-prepared team can save many pregnancies and reduce both maternal and fetal risk. The challenge of this dual-patient event is a special one to physicians.

### General References

Barron WM. Medical evaluation of the pregnant patient requiring nonobstetric surgery. Clin Perinatol 1985; 12(3):481–496.

Barron WM. The pregnant surgical patient: medical evaluation and management. Ann Intern Med 1984; 101(5): 683–691.

Brodsky JB. Anesthesia and surgery during early pregnancy and fetal outcome. Clin Obstet Gynecol 1983; 26(2): 449–457.

Delaney AG. Anesthesia in the pregnant woman. Clin Obstet Gynecol 1983; 26(4):795–800.

Duncan PG, Pope WD, Cohen MM, Greer N. Fetal risk of anesthesia and surgery during pregnancy. Anesthesiology 1986; 64(6):790–794.

James FM III. Anesthesia for nonobstetric surgery during pregnancy. Clin Obstet Gynecol 1987; 30(3):621–628.

Liu PL, Warren TM, Ostheimer GW, Weiss JB, Liu LM. Foetal monitoring in parturients undergoing surgery unrelated to pregnancy. Can Anaesth Soc J 1985; 32(5):525–532.

Slater G, Aufses A Jr. Surgical aspects of pregnancy. In: Medical, surgical and gynecologic complications of pregnancy, 3rd ed. Baltimore: Williams & Williams, 1985.

# Medical Evaluation of the Elderly Psychiatric Patient

Paul S. Aisen, M.D.

The distinction between internal medicine and psychiatry becomes blurred on a geriatric psychiatry service. Psychiatric syndromes commonly include physical symptoms; for example, depressed patients may exhibit somatic preoccupation, and psychotic patients may have somatic delusions. Conversely, physical illness in the elderly often has an atypical clinical appearance. Patients may be admitted to a psychiatric service with a diagnosis of depression, and the underlying cause (for example, adverse effect of medication, occult malignancy, chronic infection, temporal arteritis, hypothyroidism, hypercalcemia) may require careful investigation to uncover. Thus, the consulting internist plays a crucial role in diagnosis.

Internist and psychiatrist must also collaborate in therapeutic decision-making. The presence of multiple medical illnesses and polypharmacy complicate the clinical picture and the selection of treatments. The marked variability in drug response among elderly patients mandates caution and vigilance in prescribing potentially toxic drugs.

### Diagnosing the Cause of Dementia

Patients are often admitted to a geriatric psychiatry service for evaluation and management of dementia. Most of these patients have either Alzheimer's disease or multi-infarct dementia.

**Routine Evaluation.** Routine evaluation, including careful interviews of patient and family members to determine time course, identification of risk factors for dementia (family history, alcohol intake, hypertension, diabetes, hypercholesterolemia, vascular disease), physical and neurologic examination, blood studies (complete blood count, kidney and liver function tests, VDRL, thyroid function tests, vitamin $B_{12}$ and folate levels), and brain imaging (computed tomography or magnetic resonance imaging), usually leads to an accurate diagnosis. Most endocrine, metabolic, and structural causes will be readily identified.

**Clues to Unusual Causes.** However, the clinician must look for clues to unusual, potentially treatable causes of dementia. Without histopathologic confirmation, Alzheimer's disease remains a diagnosis of exclusion (1). In a heavy smoker, pulmonary disease (hypoxia, hypercarbia) must be considered. Low cardiac output may impair cognition in a patient with severe cardiomyopathy. Relatively normal liver chemistry levels may belie cirrhosis with hepatic encephalopathy.

A rapid course of dementia, particularly when it is associated with the features of systemic illness, should prompt careful consideration of unusual diagnoses, such as paraneoplastic encephalopathy secondary to occult malignancy, inflammatory diseases of the brain (sarcoidosis, granulomatous angiitis of the central nervous system, temporal arteritis), central nervous system infections (tuberculosis, tertiary syphilis, human immunodeficiency virus, Jakob-Creutzfeldt disease), and toxins (heavy metals).

### Diagnosing the Cause of Delirium

A common reason for admission to a geriatric psychiatry service is a behavioral change or cognitive deterioration in a patient with an underlying dementia. Internist and psychiatrist must collaborate in establishing a diagnosis and iden-

---

From the Department of Medicine, the Mount Sinai School of Medicine, New York, NY Address reprint requests to the author at 133 East 95th Street, New York, NY 10128.

tifying treatable causes. The clinician must be familiar with the clinical features and etiologies of delirium in the elderly.

**Delirium versus Dementia.** Delirium is an acute confusional state characterized by impairment of cognition and attention and fluctuating level of alertness (2). It is particularly common in elderly patients with underlying brain disease such as dementia. Treatable causes of delirium must be sought in all instances of apparent acceleration of a dementing illness such as Alzheimer's disease and when sudden behavioral changes occur. Features that distinguish delirium from pure dementia include

- sudden onset
- impaired level of consciousness and alertness
- marked fluctuation within the course of a day
- increased or decreased psychomotor activity
- hallucinations, often visual

Delirium may be caused by almost any drug, illness, or even physical discomfort (for example, fecal impaction) in the setting of underlying brain disease. The usual clues to the diagnosis of common medical disorders may be absent or difficult to uncover. The clinician must be a careful detective, interviewing family members and caregivers, observing and examining the patient, and initiating appropriate laboratory investigation.

**Prescription Drugs: Devising a Rational Simplification.** Prescription drugs are the most common cause of delirium in the elderly, for many reasons (2). Elderly patients often have multiple somatic complaints (such as musculoskeletal pain and gastrointestinal symptoms) for which drugs are prescribed. Impairments of vision, hearing, dexterity, and cognition contribute to dosing errors. Variability in absorption, distribution, receptor binding, metabolism, and elimination also contributes to the marked increased risk of drug toxicity in geriatric patients.

Tricyclic antidepressants, neuroleptics, anticholinergics, and benzodiazepines, all commonly used on psychiatric services, are among the drugs most likely to cause delirium. However, virtually all drugs can cause delirium in susceptible patients (3). When assessing a patient with a cognitive or behavioral deterioration, the internist must critically review all medications being taken.

**Avoiding Additional Agents.** Treatment of drug-induced delirium (misdiagnosed, for example, as dementia with psychotic features) with additional medication is an unfortunately common error. A demented patient whose behavior deteriorates while on neuroleptic treatment should not be given a higher dose of a neuroleptic agent without carefully considering the possibility of delirium. If an elderly depressed patient taking a tricyclic antidepressant develops hallucinations and a fluctuating level of alertness, the likely diagnosis is drug-induced delirium, not psychotic depression; the appropriate treatment is withdrawal of the antidepressant, not the addition of a neuroleptic agent.

**Discontinuing Drugs for Observation.** Ideally, all drugs should be discontinued for a period of observation. When the psychiatric diagnosis is unclear, discontinuing medication is usually the most helpful first step. Similarly, psychiatric diagnosis may be clarified by withdrawal of medication prescribed for nonpsychiatric indications. Anti-inflammatory agents, antihypertensives, antiarrhythmics, antianginals, and antiulcer drugs may all cause delirium, depression, mania, or anxiety (3). These agents may often be discontinued without significant risk, particularly in the hospital setting. When feasible, all drug therapy should be stopped to allow the effects of drugs to be distinguished from manifestations of a primary psychiatric disorder. Alternatively, substitution of drugs with less central nervous system activity (for example, ranitidine for cimetidine, atenolol for propranolol) may be helpful.

A major role of the consulting internist should be rational simplification of the drug regimen.

## Treating Depression

The treatment of major depression in the elderly patient with underlying medical problems requires careful consideration of the available options and their relative efficacy and toxicity. The internist may be consulted regarding the potential adverse effects of antidepressant medications, particularly in patients with underlying cardiac disease. Medical clearance is also required prior to the administration of electroconvulsive therapy.

**The Tricyclic Antidepressants.** Tricyclic antidepressants are often the first drugs considered for treatment of major depression (4–6). Because they are similar in efficacy, the choice among them should be based on relative toxicity.

*Orthostatic hypotension and anticholinergic effects,* both particularly dangerous in elderly patients, are the major concerns with drugs in this class. Anticholinergic side effects include memory impairment and frank delirium, in addition to tachycardia, dry mouth, constipation, urinary retention, and blurred vision.

As a general rule, tricyclic secondary amines (for example, nortriptyline, desipramine) are less likely to cause orthostatic hypotension or anticholinergic toxicity than tertiary amines (for example, amitriptyline, imipramine). Amitriptyline in particular should be avoided in elderly patients.

Nortriptyline is mildly sedating. It may be considered the tricyclic drug of choice in elderly patients with difficulty sleeping and is generally given as a single dose at bedtime. Desipramine tends to be activating and may be given during the day to lethargic depressed patients.

Any tricyclic drug requires careful monitoring. Patients should be checked for orthostatic hypotension before starting the drug and prior to each dose during titration. An electrocardiogram should be analyzed, with particular attention to P-R, QRS, or Q-T prolongation, prior to each increase in tricyclic dose. The clinician must be alert to potential anticholinergic toxicity, particularly cognitive impairment and, in men, urinary retention. The initial dose should be low (for example, nortriptyline, 10–25 mg at bedtime), and may be raised every three to five days. Drug levels may be helpful in determining whether a therapeutic level has been achieved, to assess the adequacy of a trial.

**Tricyclic Drugs and Underlying Heart Disease.** Treatment of patients with underlying heart disease requires particular care. Although there appears to be no significant negative inotropic effect of tricyclic antidepressants (7), patients with impaired left ventricular function are at greatly increased risk of orthostatic hypotension (8). Prolongation of cardiac conduction times, rarely of clinical importance in patients without heart disease, may lead to complete heart block in patients with preexisting bundle-branch block. Tricyclic drugs are like quinidine in their effect on ventricular arrhythmias: they tend to suppress ventricular irritability, but in some patients may be proarrhythmic, particularly when drug levels exceed the therapeutic range.

**Trazodone.** Trazodone, a newer agent useful as a soporific, is a serotonin-uptake inhibitor structurally distinct from the tricyclics. Of great importance for use with elderly patients, it is essentially without anticholinergic side effects (9).

*Orthostatic hypotension* does occur, but tends to be less significant than with the tricyclic compounds. However, trazodone has been reported to cause increased ventricular irritability or even ventricular tachycardia in a few patients with preexisting arrhythmias (10). Though the significance of these reports has been questioned (11), ventricular arrhythmia should be a relative contraindication to the use of trazodone.

**Monoamine Oxidase Inhibitors.** Monoamine oxidase (MAO) inhibitors are another important treatment option for patients who are able to comply with the necessary dietary and drug restrictions. (Cognitively impaired patients without constant supervision should not be given MAO inhibitors.) These drugs, like trazodone, have the major advantage of being free of anticholinergic effects.

*Orthostatic hypotension* is the most dangerous side effect; it may first occur several weeks after the drug is started (12). Hypertensive patients may not require other antihypertensive medication while taking MAO inhibitors. Other side effects include insomnia and weight gain. Hepatocellular damage is rare.

To avoid the possibility of hypertensive crisis, patients taking MAO inhibitors must be able to adhere to a low-tyramine diet. Sympathomimetic agents such as over-the-counter cold remedies must be avoided. Administration of meperidine is absolutely contraindicated in patients taking MAO inhibitors; fatal interactions have been reported (13). Toxic reactions have also been reported in patients taking serotonergic drugs such as fluoxetine and buspirone with MAO inhibitors (5). Patients taking MAO inhibitors are at increased risk of hypertensive and hypotensive reactions during surgery and anesthesia; these drugs should be stopped two weeks prior to elective surgery or electroconvulsive therapy.

**Lithium.** Lithium is used not only as the primary treatment of manic disorders, but also as a prophylactic agent in recurrent depression and as an adjunct with other antidepressant medications. It has a narrow therapeutic window, and its use in the elderly is complicated by

- reduced glomerular filtration rate
- fluctuations in fluid and electrolyte balance
- increased susceptibility to neurologic toxicity

The half-life of lithium in a young adult is 18 hr; in a 70-year-old patient without kidney disease it is 36 hr (14). The drug regimen should begin at low doses (75–150 mg/day), and sufficient time allowed (7–8 days) to reach steady state before levels are measured and adjustments made.

*Common side effects* of lithium which may not require discontinuation include tremor, polyuria, nausea, and mild lethargy. If necessary, tremor may be treated with low-dose propranolol;

polyuria paradoxically responds to thiazide diuretics.

*More severe toxicity* may include ataxia, confusion, seizures, or even coma. Hypothyroidism and goiter may occur; monitoring of thyroid function tests is necessary. Other side effects include weight gain, bradycardia, psoriasiform rash, blurred vision, and slurred speech.

*Renal function* should be assessed prior to lithium therapy, both as a guide to dosing and as a baseline. Renal insufficiency impairs excretion and may increase the risk of nephrotoxicity. Unfortunately, creatinine clearance may be difficult to measure in the impaired elderly patient; serum creatinine is a less reliable indicator of renal function.

*In patients with heart failure or other edematous states,* lithium therapy is complicated by the fact that lithium clearance varies with fluid and electrolyte balance. If a thiazide diuretic is added, the lithium dose should be decreased by half (guided by blood levels). Loop diuretics have somewhat less effect on lithium levels (6). Nonsteroidal anti-inflammatory drugs may increase lithium levels (15).

**Fluoxetine.** A new serotonergic antidepressant, fluoxetine, represents a major advance in safety profile. It does not commonly cause orthostatic hypotension, cardiac conduction delays, or anticholinergic effects. Side effects may include insomnia and weight loss.

*Fluoxetine in conjunction with MAO inhibitors* has been reported to cause toxic reactions, as noted above. The half-life of fluoxetine is long, so 5 weeks should elapse after stopping the drug before prescribing an MAO inhibitor (16).

**Electroconvulsive Therapy.** Electroconvulsive therapy (ECT) is the most effective treatment for severe depression (5, 16). It is indicated in patients who are psychotic, suicidal, or unwilling to eat, or who have failed to improve on trials of medication. ECT is considered by some to be less hazardous than drug treatment (17, 18).

*Risks:* The most frequent complications are cardiovascular: hypertension, hypotension, arrhythmias, and, less commonly, heart failure, ischemia, or infarction. The risk is clearly related to preexisting cardiac disease.

Elevated intracranial pressure is the only absolute contraindication to ECT, but uncontrolled angina, recent myocardial infarction or stroke, uncontrolled hypertension, and severe pulmonary disease represent relative contraindications.

The internist must assess the patient's medical status, with particular attention to cardiac status, prior to giving clearance for ECT and anesthesia. If a computed tomography or magnetic resonance imaging scan of the brain cannot be obtained prior to treatment, careful funduscopic examination is necessary to rule out increased intracranial pressure. Electroencephalogram and X-rays of the skull and spine are not essential.

Working with the psychiatrist, the consulting internist must weigh the risk of the procedure against the risk of the patient's depression. Stabilizing blood pressure and fluid status will decrease the risk. Psychotropic medications should be minimized; MAO inhibitors must be stopped two weeks prior to ECT.

Confusion and temporary memory loss are common after ECT. Irreversible cognitive impairment has not been documented. Confusion may be minimized by decreasing the frequency of treatments (to two per week or less), and by using unilateral electrode placement and low-energy stimulus waveforms (5). These measures may be particularly important in patients with baseline cognitive dysfunction.

## Choosing and Monitoring a Neuroleptic Agent

Neuroleptic medications are commonly prescribed for elderly patients, more often in an attempt to control the behavioral complications of dementia than in primary use (21).

*Extrapyramidal effects* include parkinsonism (rigidity, bradykinesia, shuffling gait, drooling, and tremor), dystonias, and akathisia. Akathisia is restlessness, jitteriness, and inability to sit still, with a vague feeling of anxiety; it may be mistakenly interpreted as a primary psychiatric symptom. Extrapyramidal symptoms may be treated with anticholinergics or amantadine. But because these drugs are so often toxic in the elderly, causing confusion and delirium, it is preferable to lower the neuroleptic dose or substitute a lower-potency neuroleptic. Akathisia may respond to benzodiazepines or beta blockers (6), but again, lowering the neuroleptic dose is preferable.

*Tardive dyskinesia* consists of a variety of involuntary movements, including lip smacking, writhing of the tongue, facial grimaces, and other tics. It is equally likely to occur with any neuroleptic, and it is particularly common in elderly patients (19). In addition to age, dementia is a risk factor, along with duration of drug exposure and maximum dose. Tardive dyskinesia often persists after the neuroleptic is stopped. There is no effective treatment. Tardive dyskinesia may be disabling and is a major concern with neuroleptic use in the elderly.

*Rare side effects* of neuroleptics most often reported with low-potency drugs include agranulocytosis, jaundice, and seizures. Thioridazine may cause cardiac conduction and repolarization abnormalities (20) and at high doses has caused irreversible retinopathy (21). Any of the neuroleptics may cause inappropriate secretion of antidiuretic hormone.

*Neuroleptic malignant syndrome* may be considered an extreme point on the spectrum of extrapyramidal side effects of antipsychotic drugs. It consists of fever, severe rigidity, altered consciousness, and elevated muscle enzymes. This syndrome carries a significant mortality rate and must be considered whenever fever occurs with extrapyramidal symptoms in patients taking neuroleptics. Treatment consists of discontinuing antipsychotic medication and providing supportive care. Medications that may be useful in managing neuroleptic malignant syndrome include dantrolene and bromocriptine (22). If antipsychotic treatment is necessary in a patient who has recovered from neuroleptic malignant syndrome, a careful trial of a low-potency drug such as thioridazine may be considered (23).

*Clozapine* is a new, unique neuroleptic essentially free of extrapyramidal effects (24). It is currently indicated for treatment-resistant schizophrenia, but it may be useful in patients with Parkinson's disease or extrapyramidal toxicity from traditional antipsychotics. Clozapine commonly causes sedation and hypotension. Its most serious side effect is agranulocytosis, occurring in about 1% of patients treated for one year (25). Weekly monitoring of blood counts is required during clozapine treatment.

## Choosing and Monitoring an Antianxiety Agent

Benzodiazepines are prescribed for elderly patients to control behavioral abnormalities complicating dementia and to induce sleep, in addition to their indication for anxiety disorders and other functional psychiatric illnesses. The half-life of these drugs is generally long, and most have active metabolites. For example, diazepam and its active metabolites have a mean half-life of 20 hr in young adults and 90 hr in 80-year-old patients (26). Thus, with daily use, clinical effects are likely to build up over many days.

- Lorazepam, with a shorter half-life, no active metabolites, and little change in metabolism with aging, is a good choice when a benzodiazepine is indicated for an elderly patient.
- Triazolam, with a fast onset and short half-life, is appropriate for sleep induction.

**Side Effects and Falls.** Side effects of benzodiazepines, which include lethargy, memory and cognitive impairment, weakness, and ataxia, are problematic in elderly patients. These drugs may increase the risk of falls and hip fractures, especially in patients with underlying cognitive or motor dysfunction (27). Withdrawal symptoms, which in severe cases may include delirium and convulsions, are most prominent with short-acting benzodiazepines; weaning is essential when use has been long term.

**Alternatives.** Alternative drugs should be considered for elderly patients.

- For sleep induction, trazodone (particularly when there may be a component of depression) or chloral hydrate may be preferable.
- Buspirone, a new partial serotonergic agonist, is a nonsedating and non-habit-forming anxiolytic (5). Although it has a slow onset, requiring up to a week for apparent benefit, its safety makes it a good option for the treatment of anxiety in elderly patients.

## Summary

Elderly patients with psychiatric disorders present a formidable clinical challenge. Medical and psychiatric disorders coexist and mimic each other, blurring diagnostic criteria. Accurate diagnosis is essential and can best be achieved by cooperation between internist and psychiatrist. Most treatments carry substantial risk, which must be minimized by careful medical assessment and optimal selection among options. Error is unavoidable, so clinicians must continually analyze and question their diagnostic and therapeutic decisions.

### References

1. Brumback RA, Leech RW, Carella J, Miner GD. Dementia: the importance of clinical evaluation, autopsy confirmation, and research. J Oklahoma S Med Soc 1990; 83:109–118.
2. Lipowski ZJ. Delirium in the elderly patient. N Engl J Med 1989; 320:578–582.
3. Drugs that may cause psychiatric symptoms. Med Lett 1989; 31:113–118.
4. Jarvik MA. Antidepressant therapy for the geriatric patient. J Clin Psychopharmacol 1981; 1:55S–61S.

5. Jenike MA. Geriatric psychiatry and psychopharmacology: a clinical approach. Chicago: Year Book Medical Publishers, 1989.
6. Davidson J. The pharmacologic treatment of psychiatric disorders in the elderly. In: Busse EW, Blazer DG, eds. Geriatric psychiatry. Washington, DC: American Psychiatric Press, 1989.
7. Veith RC, Raskind MA, Caldwell JH, et al. Cardiovascular effects of tricyclic antidepressants. N Engl J Med 1982; 306:954–959.
8. Glassman AH, Johnson LL, Giardina EGV, et al. The use of imipramine in depressed patients with congestive heart failure. JAMA 1983; 250:1997–2001.
9. Branconnier RJ, Cole JO. Effects of acute administration of trazodone and amitriptyline on cognition, cardiovascular function and salivation in the normal geriatric subject. J Clin Psychopharmacol 1981; 1:82S–88S.
10. Janowski D, Curtis G, Zisook S, et al. Ventricular arrhythmia possibly aggravated by trazodone. Am J Psychiatry 1983; 140:796–797.
11. Himmelhoch JM, Schechtman K, Auchenbach R. The role of trazodone in the treatment of depressed cardiac patients. Psychopathology 1984; 17(suppl 2):51–63.
12. Kronig MH, Roose SP, Walsh BT, et al. Blood pressure effects of phenelzine. J Clin Psychopharmacol 1983; 3:307–310.
13. Meyer D, Halfin V. Toxicity secondary to meperidine in patients on monoamine oxidase inhibitors: a case report and critical review. J Clin Psychopharmacol 1981; 1:319–326.
14. Jefferson JW. Lithium and affective disorders in the elderly. Compr Psychiatry 1983; 24:166–178.
15. Reimann IW, Frolick JC. Effects of diclofenac on lithium kinetics. Clin Pharmacol Ther 1981; 30:348–352.
16. Blazer DG. Affective disorders in late life. In: Busse EW, Blazer DG, eds. Geriatric psychiatry. Washington, DC: American Psychiatric Press, 1989.
17. Fink M. Convulsive therapy: theory and practice. New York: Raven Press, 1979.
18. Weiner RD. The role of electroconvulsive therapy in the treatment of depression in the elderly. J Am Geriatr Soc 1982; 30:710–712.
19. Jenike MA. Tardive dyskinesia: special risk in the elderly. J Am Geriatr Soc 1983; 31:71–73.
20. Gelenberg AJ. Treating the outpatient schizophrenic. Postgrad Med 1978; 64:48–55.
21. Baldessarini RJ. Chemotherapy in psychiatry. Cambridge, MA: Harvard University Press, 1977.
22. Rosenberg MR, Green M. Neuroleptic malignant syndrome: review of response to therapy. Arch Intern Med 1989; 149:1927–1931.
23. Gelenberg AJ. Return to NMS. Biol Ther Psychiatry 1986; 9:17–20.
24. Lieberman JA, Kane JM, Johns CA. Clozapine: guidelines for clinical management. J Clin Psychiatry 1989; 50:329–338.
25. Clozapine for severe schizophrenia. FDA Drug Bull 1990; 20:9.
26. Rosenbaum J. Anxiety. In: Lazare A, ed. Outpatient psychiatry. Baltimore: Williams & Wilkins, 1979.
27. Ray WA, Griffen MR, Schaffner W, et al. Psychotropic drug use and the risk of hip fracture. N Engl J Med 1987; 316:363–369.

*Brief Communication*

# Physician-Nurse Collegiality in the Medical School Curriculum:

## Exploratory Workshop and Student Questionnaire

Patricia Lynch Anvaripour, MPA, RN, CNAA, Linbania Jacobson, EdD, RN, Joyce Schweiger, MS, RN, and Gail Kuhn Weissman, EdD, RN, FAAN

Abstract

Medical education is the key to improvements in physician-nurse relationships and communication. The results of a workshop for second-year medical students on collegiality organized by The Mount Sinai Hospital Department of Nursing and the Mount Sinai School of Medicine are presented. The workshop follow-up roundtable discussion is summarized and the results of a survey of medical students' perception of professional nursing are presented.

Concern with the widespread difficulties of recruiting and retaining registered professional nurses has forced close scrutiny and study of the reasons, factors, and forces contributing to this national crisis. Although the current shortage is widely recognized as due primarily to an increase in demand as opposed to a concentration of supply, it is acknowledged that limited career advancement opportunities, lack of collegial relationships between physicians and nurses, and minimal utilization of nurses' specialized knowledge have an impact on recruitment into the profession (1–4).

For example, the final report issued by the Secretary's Commission on Nursing, a public advisory panel established to develop recommendations on how the public and private sectors can work together to implement immediate and long-range solutions for enhancing the adequacy of the supply of nurses, stated that "the medical profession should recognize the appropriate clinical decision making authority of nurses in relationship to other health care professionals, foster communication and collaboration among the health care team, and [recognize that] close cooperation and mutual respect between nursing and medicine is essential" (5). A study by Knaus, Draper, Wagner, and Zimmerman (6) concluded that the quality of the relationship between physicians and nurses in intensive care units is the vital element in lower patient mortality rates. The Mount Sinai Medical Center is developing a new initiative in physician-nurse relationships that begins early in the education of medical students.

In 1987, Gail Kuhn Weissman, the Vice President of Nursing at The Mount Sinai Medical Center established a Physician Advisory Board to identify trends in medical care that would have an impact on nursing practice and quality of patient care, and to explore ideas on how to best meet these challenges at Mount Sinai (7). The Advisory Board includes physicians from the major clinical services, the Associated Alumni of Mount Sinai and the Alumni Association of Attending Staff, and a member of the Board of Trustees. Physician interpersonal communications was among the issues defined by the Advi-

---

From the Department of Nursing, The Mount Sinai Hospital. Address reprint requests to Patricia Lynch Anvaripour, MPA, RN, CNAA, Department of Nursing, Box 1157, The Mount Sinai Medical Center, One Gustave L. Levy Place, New York, NY 10029.

sory Board as in need of immediate enhancement. The Advisory Board developed and recommended to the dean of The Mount Sinai School of Medicine and to individual departmental chairmen the adoption of a plan to improve the communication skills of all health professionals.

### Workshop on Physician-Nurse Collegiality

As a result, early in 1990, the Department of Nursing and the School of Medicine developed a workshop on physician-nurse collegiality for the course Clinical Assessment of Patients, so that second-year students could learn about (a) differentiated practices of nurses and (b) how to improve communication with other professional caregivers. Close to half of the second-year class took advantage of this innovative learning experience. The 60 students were divided into three groups of 20 each, for participation in one two-hour Friday afternoon workshop.

The workshop was planned to ensure a learning experience that included the clinical setting. Each workshop began with a one-hour panel presentation in which a staff nurse, a senior clinical nurse, a nurse clinician, a nurse educator, and a nurse clinical supervisor each briefly described their educational and professional background, specific role and responsibility in the clinical setting, and day-to-day interactions with physicians.

The nurses explained the nursing process: assessment, diagnosis, care planning, treatment, and evaluation. They also described their undergraduate scientific preparations in biology, chemistry, physics, anatomy, and physiology. Additionally, nurses with advanced preparation in a specialty area talked about the rigorous process of the education they received and the examinations they must pass to be recognized as specialists. A question-and-answer period followed.

In the second hour, each student paired off with a staff nurse mentor and visited the student's preselected clinical setting: neonatology, cardiology, obstetrics, pediatrics, or neurology. During this time, students learned about how nurses and physicians on these units work together in a collegial fashion daily. For example, during scheduled team meetings, nurses and physicians share in the planning and setting of patient care goals, in decision making, and in solving problems. Teaching rounds provide an additional opportunity to reinforce collegial patient-care behaviors. Educational conferences provide yet another opportunity for nurses and physicians to share their specialized knowledge with each other. On the unit, student and nurse reviewed a patient chart and briefly interviewed the patient about his or her perception of the nurse-physician relationship.

### Evaluation

**Methodology.** The purpose of the evaluation was to determine:

1. What did participating medical students and nurses learn about each other's profession as a result of the workshop?
2. How can future workshops be improved?
3. What did participating and nonparticipating medical students know about professional nursing and nurses' roles and skills?

To answer these questions, both qualitative and quantitative data collection methods were used. An end-of-workshop focus group session with select participants representing medical students, nurse-mentors, panelists, program planners, and others was conducted. Focus group participants discussed trigger questions that had been pre-

TABLE 1

*Questionnaires on Physician-Nurse Collegiality Returned by Mount Sinai School of Medicine Students*

| | No. | % | No. of questionnaires mailed |
|---|---|---|---|
| **Medical school year** | | | |
| First | 12 | 20 | 118 |
| Second | 22 | 36 | 117 |
| Third | 10 | 16 | 118 |
| Fourth | 17 | 28 | 131 |
| Total | 61 | 100 | 485 |
| **Sex** | | | |
| Male | 30 | 49 | |
| Female | 31 | 51 | |
| Total | 61 | 100 | |
| **Age Group** | | | |
| 20–24 | 26 | 43 | |
| 25–29 | 25 | 41 | |
| 30–34 | 7 | 11 | |
| 35–39 | 3 | 5 | |
| Total | 61 | 100 | |
| **Participated in workshop*** | | | |
| No† | 7 | 32 | |
| Yes | 15 | 68 | |
| Total | 22 | 100 | |
| **Reason for not participating†** | | | |
| Not interested | 1 | | |
| Not informed | 1 | | |
| Other commitments | 5 | | |

* Only second-year students were eligible to participate in workshop and answered the question.

† The 7 nonparticipants in the workshop.

pared in advance by the group leader. A summary of this discussion is presented below, along with recommendations made by the combined group.

**Summary of Medical Students' Comments.** Students recognize nursing as an autonomous professional body that has the greatest contact with patients and information about patient care management. Thus, students want to benefit from nurses' vast store of patient-care knowledge. They indicated that fourth-year medical students and housestaff officers would also like to ask nurses about patient care, for example about their assessment and findings, but are afraid to do so. Students also indicated that they perceive a barrier between nurses and physicians and that physicians do not listen and are often condescending to nurses.

Additionally, students indicated that they feel frustrated about their limited patient contact. They think that nurses think students have no clinical skills to offer, and that nurses lack understanding of students' clinical training. Students expressed surprise and amazement at a nurse's comment about a medical student: "Can't even draw blood." They also felt they have no opportunity to develop a relationship with nurses and do not receive any orientation to patient care units, or to the roles of the different health team members. Residents, who also know very little about different health provider roles, orient them to the units. The collegiality workshop was a very nice, but very brief, exposure to "nurses in action."

**Summary of Nurse-Mentors' Comments.** "Students are like sponges, eager and thirsty for knowledge, quick to soak up any patient care information." Some nurses understand that students, at this point in their training, have limited clinical experience, and that all student orders must be countersigned by a house officer. However, nurses know very little else about the students' medical training.

Nurses felt that some of their peers were cynical about the long-term effects of the workshop: "Not going to make any difference in their behavior once they are physicians."

**Summary of Recommendations Made by Medical Students and Nurse-Mentors.** Physician-nurse collegiality needs to be stressed and integrated throughout all four years of medical school and to be continued into the postgraduate years.

The topic should be initiated during the first three weeks of medical school, and included as a workshop during the course on stress reduction and communications.

The topic should be reintroduced during the second year, when students are enrolled in the physical assessment course.

Follow-up workshops should include physician-nurse case presentations in the third and fourth years. Future workshop lengths should be a minimum of one-half to one full day.

The physician-nurse collegiality concept should be introduced to housestaff officers during their orientation. They need to learn how to approach nurses for patient information. Nurses can educate new house officers by introducing themselves when the physicians first arrive on

TABLE 2

*Skills and Abilities for a Nurse to Have, Ranked from Most Important (10) to Least Important (0)*

| | Rating |
|---|---|
| **Very important** | |
| Reliability | 9.62 |
| Sensitivity and compassion | 9.12 |
| Friendly and courteous | 8.56 |
| Good technical skills | 8.38 |
| Intelligence | 8.35 |
| Handle many things at one time | 8.34 |
| Good decision maker | 7.98 |
| Take control of situations | 7.52 |
| Good problem solver | 7.44 |
| Work independently | 7.08 |
| **Least important** | |

TABLE 3

*Functions and Duties of a Nurse, Ranked from Most Important (1.0) to Least Important (3.0)*

| | Rating |
|---|---|
| **Major functions & duties (rank 1.0–1.99)** | |
| Gives medications to patients | 1.0 |
| Talks with physicians about patient's needs | 1.12 |
| Takes patient's blood pressure and pulse | 1.36 |
| Provides support for a patient's family | 1.55 |
| Fills out paperwork | 1.55 |
| Starts IV's | 1.64 |
| Keeps patients informed about their condition | 1.80 |
| Provides information to patients about medications | 1.82 |
| Plans for the care of patients | 1.84 |
| **Minor functions & duties (rank 2.0–2.99)** | |
| Transports patients for tests | 2.0 |
| Runs errands, such as takes samples to the labs | 2.53 |
| Removes stitches after an operation | 2.63 |
| Delivers babies | 2.76 |
| Gives anesthesia to patients | 2.81 |
| **Not part of a nurse's job (rank 3.0)** | |

TABLE 4
*Opinion about the Nursing Profession, Ranked from Agree Strongly (1.0) to Disagree Strongly (4.0)*

| | Rating |
|---|---|
| Agree strongly<br>(rank 1.0–1.99) | |
| It is important for nurses to have a good knowledge of medications | 1.33 |
| There are many specialized fields within the general field of nursing | 1.37 |
| Nurses usually work long hours | 1.42 |
| Nursing is a challenging career | 1.53 |
| Nursing jobs are easy to find | 1.56 |
| It is important for nurses to have a good knowledge of diseases and disorders | 1.56 |
| Nursing is an emotionally rewarding career | 1.81 |
| Agree somewhat<br>(rank 2.0–2.99) | |
| It is easy to get into nursing school | 2.06 |
| Nurses require good technical skills to operate medical equipment | 2.07 |
| Nurses have flexibility in their schedules | 2.25 |
| It takes a lot of scientific training to become a nurse | 2.35 |
| There are opportunities for career advancement in the field of nursing | 2.44 |
| Generally, nurses are fairly paid for the work they do | 2.84 |
| Disagree somewhat<br>(rank 3.0–3.99) | |
| Anyone who cares about people could be a nurse | 3.33 |
| Disagree strongly<br>(rank 4.0) | |

TABLE 5
*Recommend Registration in Workshop on Physician-Nurse Collegiality to Other Medical Students*

| | No. | % |
|---|---|---|
| Extremely likely | 15 | 25 |
| Very likely | 25 | 41 |
| Somewhat likely | 13 | 21 |
| Not too likely | 7 | 11 |
| Not at all | 1 | 1 |
| No answer | 1 | 1 |
| Total | 61 | 100 |

the units and explaining how nurses can be helpful.

## Survey of Medical Students' Image of Professional Nurses

Near the end of the academic year a survey questionnaire was mailed to a total of 485 Mount Sinai medical students. The instrument had been developed for use by the American Nurses Association's National Commission on Nursing Implementation Project. The instrument measures consumers' perception of the nursing profession and attitudes toward nursing as a career. The questionnaire consists of three rating scales designed to obtain information on:

- Skills and qualifications important to nurses
- Functions and duties of a nurse's job
- Opinions about the nursing profession
- Demographics

Six weeks after mailing the questionnaire, a total of 61 surveys were returned—obviously not a representative sample of the population, but nontheless a beginning step in learning about the views of a highly self-selected group of medical students (Table 1). On the scale that rates the relative importance of certain nursing skills and abilities on a scale of 10 to 0, 10 being most important, all 14 items were considered important (Table 2). Another scale rates major and minor nursing functions and duties (Table 3). All 14 listed functions were considered a part of a nurse's job, and a majority of them (9/14) were rated as a major part. And the third rates agree and disagree statements about the nursing profession (Table 4). Respondents agreed with all but one statement about the nursing profession.

That 53/61 respondents would recommend registration in a workshop on physician-nurse collegiality to other medical students (Table 5) is a positive sign of the students' interest in and concern with this extremely important issue.

## The Future

Preliminary discussions are taking place concerning integrating physician-nurse collegiality throughout the undergraduate and postgraduate years of medical education. Among other initiatives fostering and supporting physician-nurse collegiality is a study currently being conducted to determine the current status of physician-nurse collaboration on inpatient care units at the medical center.

### References

1. Findings and recommendations of HANY's Nursing Study. Albany, NY: Hosp Assoc NYS, August 1988.
2. Guidelines for Establishing Joint or Collaborative Practice in Hospitals. Chicago, Illinois: The National Joint Practice Commission, 1981.
3. Moss J, Totten M. The nursing shortage and the hospital governing board. Trustee 1988; November 8–9.
4. Robert M, Minnick A, Ginzberg E, Curran C. A Commonweath Fund paper: What to do about the nursing shortage. New York, NY: The Commonwealth Fund, 1989.
5. Secretary's Commission on Nursing. Washington, DC: Department of Health and Human Services; December 1988.
6. Knaus W, Draper E, Wagner D, Zimmerman J. An evaluation of outcomes from intensive care in major medical centers. Ann Int Med 1986; 104(3):410–418.
7. Charge to the Vice President of Nursing's Physician Advisory Board. The Mount Sinai Medical Center, 1987.

# Instructions for Authors

THE MOUNT SINAI JOURNAL OF MEDICINE, established in 1934 as *The Journal of the Mount Sinai Hospital,* is a peer-reviewed general medical journal. It is indexed in *Index Medicus, Hospital Literature Index,* and *International Nursing Index,* and is abstracted in *Excerpta Medica, Chemical Abstracts,* and *Biological Abstracts.* Original articles, theoretical explorations, basic science articles, reviews, reports of clinical or research experience, articles on history or other subjects affecting the practice of medicine, short notes in any specialty, and letters are welcome, as are articles by medical students. The *Journal* is a participant in the Agreement on Uniform Requirements for Manuscripts Submitted to Biomedical Journals.

**Communications.** Address correspondence on manuscript submission and review to: Managing Editor, *The Mount Sinai Journal of Medicine,* Box 1094, 19 E. 98th Street, 4D, New York, NY 10029-6574. Telephone: (212) 241-6108.

**Copyright.** The publisher reserves copyright and renewal rights on all published material, and such material may not be published elsewhere without written permission of the publisher. When submitting a manuscript, include the following assignment of copyright in your cover letter: **In consideration of The Mount Sinai Journal of Medicine (MSJM) taking action and reviewing and editing my (our) submission, the author(s) undersigned hereby transfer(s), assign(s), or otherwise convey(s) all copyright ownership to The Mount Sinai Journal of Medicine in the event that such work is published by MSJM.** Each author must sign the statement.

**Manuscript Preparation.** Submit manuscripts in triplicate, including three sets of illustrations. Manuscripts must be typewritten, **double-spaced throughout (including references, legends, and tables),** on one side of 8½ × 11-in. white bond paper, with generous margins. Number pages consecutively. Begin a new page for title page, abstract, first page of text, references, each table, and legends. **Computer-generated typescripts** must be double-spaced and must be easy for editors and typesetters to read. **Computer tapes or disks:** consult managing editor.

**Title Page.** On the title page, on separate lines (double-spaced), include (1) title of article, phrased as concisely as possible; (2) name of each author, including first name and degree; (3) name and address of departments and institutions from which the work originated; (4) acknowledgments of sources of support; (5) name, place, and date of any conferences at which paper has been delivered; (6) the statement **Address reprint requests to . . . ,** giving full name and address, with zip code, of the appropriate author.

**Abstract.** Original articles should include an abstract, which should be a single paragraph not exceeding 150 words. State the purposes; basic approach or methods; main findings; principal conclusions. Emphasize new and important aspects. When an abstract is provided, a summary at the end is not necessary. In literature reviews and other articles which do not begin with an abstract, the manuscript should conclude with a short summary.

**Key Words.** Supply 3 to 10 key words or phrases at the bottom of the title page, for use in indexing the article. Use Medical Subject Headings from *Index Medicus.*

**Writing Style and Abbreviations.** Authors should aim for conciseness and clarity without repetition. Avoid medical jargon, abbreviated phrasing, and obscure or newly coined abbreviations. Define all abbreviations on first use, except those for standard, universally recognized units of measurement and statistical terms. Spell out such terms as white blood cell, hematocrit, superior vena cava.

**Units of Measurement.** Use S1 units, giving range or interval of normal values; more familiar units may be used in addition.

**Editing.** All papers will be edited to enhance conciseness and clarity without altering meaning and to insure conformity with journal style. The author is responsible for all statements in the article.

**Illustrations.** Photographs, line drawings, graphs, and charts should increase understanding of the text. Line drawings, graphs, and charts should be professionally prepared. All illustrations should be submitted in triplicate as sharp, high-contrast glossy prints. Photomicrographs must be 5⅛ in. wide. On the back of each print affix a gum label with the number of the figure (numbered consecutively in arabic numerals as cited in text), title of manuscript, name of senior author, and the word "top" with an arrow indicating top edge. Illustrations which are to appear together should be mounted accordingly as plates and should correspond to each other in size. Protect the prints by enclosing them in heavy cardboard; do not use paper clips. Photographs of patients can be published only if a copy of the permission is submitted with the photograph. Eliminate names of patients and hospital numbers from x-rays.

**Legends.** Legends should not duplicate the text. Begin legends on a separate sheet of paper, double spaced. Each legend should be numbered to match the illustration it accompanies. Illustrations mounted as plates to appear together should be accompanied by a single legend identifying the separate elements by position ("upper left," and so on) or by letters. If letters are used in the legend, a corresponding letter must appear on the print itself. Identify all abbreviations used in a figure. Indicate magnification and stain for photomicrographs.

**Tables.** Each table should be typed on a separate page, double spaced. Number tables consecutively as they appear in text, using arabic numerals ("Table 1"). Give each table a brief title (for example, "Effect of DMSO on Rats") and type it on a separate line. If abbreviations or symbols are used in the table, identify them in a footnote.

**References.** When citing references in your text, the first work cited is numbered 1, and succeeding references are numbered in sequence; enclose the citation number in parentheses and place it before any punctuation: **Cytomegalovirus (1), steroids (2), and recreational drugs (3) have all been implicated (4).** The reference list includes only works cited in the text, numbered in the order in which they are cited. Type, double spaced, following the general arrangement and punctuation style in the examples below (journal title abbreviations conform to *Index Medicus* style):

**1. Keskey TS, Suarez M, Gleicher N, Friberg J, Levy HB. *Chlamydia trachomatis* infection in sexually abused children. Mt Sinai J Med 1987;53:129–134.**

**2. Butler RN, Lewis MI. Love and sex after 40: a guide for men and women for their mid and later years. New York: Harper & Row, 1986, 201.**

**Proofs.** Proofs are sent to the corresponding author from the printer. Stylistic changes which do not alter meaning should not be made at this stage. Because of increased printing charges, the cost of excessive author's alterations beyond routine correction of typographical errors and major errors in data may be billed to the author. Proofs should be corrected and returned to the editorial office within 48 hours.

**Reprint Orders.** A reprint order form is sent to the corresponding author with proofs and is returned to the editorial office.

# THE
# MOUNT SINAI JOURNAL
# OF MEDICINE

## Forthcoming:

## Grand Rounds

**Bone Marrow Transplant**
Blanche Alter, MD
The Mount Sinai Hospital

**Is There A Future for Physicians in America?**
Uwe E. Reinhardt, MD
Princeton University

**Age-Associated Changes in Cardiovascular Response to Exercise**
Myron Weisfeldt, MD
Johns Hopkins Hospital, Baltimore

LEVY LIBRARY, MT. SINAI MEDICAL CENTER

## Theme Issues

**MARCH 1991**
**Lumbar Disk Disease: Controversies in Neurosurgery**
*Editors:* Martin Savitz, MD, and Leonard Malis, MD

**MAY 1991**
**Pain**
*Editor:* Allan Reed

## Available now:

**Social Implications of Health Care: Essays in Honor of Helen Rehr**
Gary Rosenberg and Bess Dana, editors
volume 56, no. 6, November 1989
80 pages $15

**Sexually Transmitted Diseases**
Edward J. Bottone, editor
volume 57, no. 4, September 1990
88 pages $15

**Inflammatory Bowel Disease Therapies for the 1990s Scientific Basis and Clinical Exp**
David B. Sachar, editor
volume 57, no. 5, October 1990
64 pages $15

*Subscriptions for 1991: $60 individuals, U.S. and Canada, and $65 all libraries; $65 individuals outside U.S. and Canada.*

*To order subscriptions or copies of back issues ($15 each), please send check, payable to The Mount Sinai Journal of Medicine, to the Journal at Box 1094, One Gustave L. Le New York, NY 10029-6574; counter sales at Suite 4D, 19 E. 98th Street, New York, NY (phone: 212 241-6108; FAX, 212 831-1816).*

THE
MOUNT SINAI SCHOOL OF MEDICINE
OF THE CITY UNIVERSITY OF NEW YORK
OUNT SINAI JOURNAL
F MEDICINE

UME 58 NUMBER 2 MARCH 1991

# Lumbar Disk Disease
# Controversies in Neurosurgery

**Leonard I. Malis and Martin H. Savitz, editors**

From the Department of Neurosurgery, Mount Sinai School of Medicine

Veterans Administration Medical Center
Bronx, New York

BETH
ISRAEL
MEDICAL
CENTER

Where there's smoke...there may be bronchitis
Ceclor®
cefaclor
Pulvules® 250 mg

"Recent research has delineated early, more subtle changes in lung and immune functions. These alterations directly predispose smokers to respiratory tract infection."

*Am Fam Phys* 1987;36:133-140

*Established therapy for today's patients*

For respiratory tract infections due to susceptible strains of indicated organisms

**Brief Summary.**

**Consult the package literature for prescribing information.**

**Indication:** Lower respiratory infections, including pneumonia, caused by *Streptococcus pneumoniae, Haemophilus influenzae,* and *Streptococcus pyogenes* (group A $\beta$-hemolytic streptococci).

**Contraindication:** Known allergy to cephalosporins.

**Warnings:** CECLOR SHOULD BE ADMINISTERED CAUTIOUSLY TO PENICILLIN-SENSITIVE PATIENTS. PENICILLINS AND CEPHALOSPORINS SHOW PARTIAL CROSS-ALLERGENICITY. POSSIBLE REACTIONS INCLUDE ANAPHYLAXIS.

Administer cautiously to allergic patients.

Pseudomembranous colitis has been reported with virtually all broad-spectrum antibiotics. It must be considered in differential diagnosis of antibiotic-associated diarrhea. Colon flora is altered by broad-spectrum antibiotic treatment, possibly resulting in antibiotic-associated colitis.

**Precautions:**

- Discontinue Ceclor in the event of allergic reactions to it.
- Prolonged use may result in overgrowth of nonsusceptible organisms.
- Positive direct Coombs' tests have been reported during treatment with cephalosporins.
- Ceclor should be administered with caution in the presence of markedly impaired renal function. Although dosage adjustments in moderate to severe renal impairment are usually not required, careful clinical observation and laboratory studies should be made.
- Broad-spectrum antibiotics should be prescribed with caution in individuals with a history of gastrointestinal disease, particularly colitis.
- Safety and effectiveness have not been determined in pregnancy, lactation, and infants less than one month old. Ceclor penetrates mother's milk. Exercise caution in prescribing for these patients.

**Adverse Reactions:** (percentage of patients)

Therapy-related adverse reactions are uncommon. Those reported include:

- Hypersensitivity reactions have been reported in about 1.5% of patients and include morbilliform eruptions (1 in 100). Pruritus, urticaria, and positive Coombs' tests each occur in less than 1 in 200 patients. Cases of **serum-sickness-like** reactions have been reported with the use of Ceclor. These are characterized by findings of erythema multiforme, rashes, and other skin manifestations accompanied by arthritis/arthralgia, with or without fever, and differ from classic serum sickness in that there is infrequently associated lymphadenopathy and proteinuria, no circulating immune complexes, and no evidence to date of sequelae of the reaction. While further investigation is ongoing, **serum-sickness-like** reactions appear to be due to hypersensitivity and more often occur during or following a second (or subsequent) course of therapy with Ceclor. Such reactions have been reported more frequently in children than in adults with an overall occurrence ranging from 1 in 200 (0.5%) in one focused trial to 2 in 8,346 (0.024%) in overall clinical trials (with an incidence in children in clinical trials of 0.055%) to 1 in 38,000 (0.003%) in spontaneous event reports. Signs and symptoms usually occur a few days after initiation of therapy and subside within a few days after cessation of therapy; occasionally these reactions have resulted in hospitalization, usually of short duration (median hospitalization = two to three days, based on postmarketing surveillance studies). In those requiring hospitalization, the symptoms have ranged from mild to severe at the time of admission with more of the severe reactions occurring in children. Antihistamines and glucocorticoids appear to enhance resolution of the signs and symptoms. No serious sequelae have been reported.
- Stevens-Johnson syndrome, toxic epidermal necrolysis, and anaphylaxis have been reported rarely. Anaphylaxis may be more common in patients with a history of penicillin allergy.
- Gastrointestinal (mostly diarrhea): 2.5%
- Symptoms of pseudomembranous colitis may appear either during or after antibiotic treatment.
- As with some penicillins and some other cephalosporins, transient hepatitis and cholestatic jaundice have been reported rarely.
- Rarely, reversible hyperactivity, nervousness, insomnia, confusion, hypertonia, dizziness, and somnolence have been reported.
- Other: eosinophilia, 2%; genital pruritus or vaginitis, less than 1% and, rarely, thrombocytopenia and reversible interstitial nephritis.

Abnormalities in laboratory results of uncertain etiology.

- Slight elevations in hepatic enzymes.
- Transient lymphocytosis, leukopenia, and, rarely, hemolytic anemia and reversible neutropenia.
- Rare reports of increased prothrombin time with or without clinical bleeding in patients receiving Ceclor and Coumadin concomitantly.
- Abnormal urinalysis; elevations in BUN or serum creatinine.
- Positive direct Coombs' test.
- False-positive tests for urinary glucose with Benedict's or Fehling's solution and Clinitest® tablets but not with Tes-Tape® (glucose enzymatic test strip, Lilly).

PA 8791 AMP [021490LRI]

*Additional information available to the profession on request from Eli Lilly and Company, Indianapolis, Indiana 46285.*

Lilly

**Eli Lilly Industries, Inc**
Carolina, Puerto Rico 00630
A Subsidiary of Eli Lilly and Company
Indianapolis, Indiana 46285

CR-0525-B-049333 © 1990, ELI LILLY AND COMPANY

HE
OUNT SINAI JOURNAL
F MEDICINE

Volume 58
Number 2
March 1991

# CONTENTS

## LUMBAR DISK DISEASE: CONTROVERSIES IN NEUROSURGERY

**Editorial** *Martin H. Savitz* .......... 95

**Conservative Management of Low Back Pain** *Richard A. Frieden* .......... 97

**Management of Low Back Pain by Analgesics and Adjuvant Drugs** *Joseph R. DiPalma and G. John DiGregorio* .......... 101

**Arthritis of the Lumbar Spine** *Gary V. Gordon* .......... 109

**Low Back Pain in Children** *Bruce R. Rosenblum and Allen S. Rothman* .......... 115

**Lumbar Stenosis** *Leonard I. Malis* .......... 121

**Synovial Cysts of the Lumbar Facet Joint** *Ved P. Sachdev, Martin H. Savitz, Antoine I. Hindi, and Howard B. Goldstein* .......... 125

**Preoperative and Postoperative Neuroradiologic Evaluation of Lumbar Disk Herniation** *Robert Tash, Richard C. Strauss, and Martin H. Savitz* .......... 129

**Role of Laminectomy in Lumbar Disk Herniation** *Charles A. Fager* .......... 133

**Microdiskectomy—Myth, Mania, or Milestone? An 18-Year Surgical Adventure** *Robert W. Williams* .......... 139

**Lumbar Diskectomy under the Operating Microscope** *Ved P. Sachdev* .......... 147

**Automated Percutaneous Lumbar Diskectomy** *Gary Onik* .......... 151

**Arthroscopic Microdiskectomy** *Parviz Kambin* .......... 159

**Minilaminotomy as an Alternative to Laminectomy or Microdiskectomy: Ten Years' Experience** *Martin H. Savitz* .......... 165

**Spinal Instrumentation for Degenerative Disease of the Lumbar Spine** *Michael J. Harrison and Narayan Sundaresan* .......... 169

**Percutaneous Radiofrequency Rhizotomy of the Lumbar Facets: Ten Years' Experience** *Martin H. Savitz* .......... 177

**Risk of Wound Infection in Lumbar Disk Surgery** *Sean I. Savitz, Lillian V. Lee, and Howard B. Goldstein* .......... 179

**Causes of Failure of Surgery on the Lumbar Spine: Ten-Year Follow-up** *Charles V. Burton* .......... 183

# THE MOUNT SINAI JOURNAL OF MEDICINE

The Mount Sinai Journal of Medicine is published by The Mount Sinai Medical Center of New York and has the following affiliates: Beth Israel Medical Center, New York; Bronx Veterans Administration Medical Center, New York; and Elmhurst Hospital Center, New York.

**Editor-in-Chief**

Sherman Kupfer, M.D.

**Editors Emeriti**

David A. Dreiling, M.D. Lester R. Tuchman, M.D.

**Associate Editors**

Harriet S. Gilbert, M.D. Julius Wolf, M.D.

**Managing Editor**

Claire Sotnick

**Assistant Editors**

David H. Bechhofer, Ph.D.
Constanin A. Bona, M.D., Ph.D.
Edward J. Bottone, Ph.D.
Jurgen Brosius, Ph.D.
Steven T. Brower, M.D.
Lewis Burrows, M.D.
Joseph S. Eisenman, Ph.D.
Richard A. Frieden, M.D.
Steven Fruchtman, M.D.
Paul L. Gilbert, M.D.
James H. Godbold, Ph.D.
Richard S. Haber, M.D.
Noam Harpaz, M.D.
Dennis P. Healy, Ph.D.
Barry W. Jaffin, M.D.
Andrew S. Kaplan, D.D.S.
Samuel Kenan, M.D.
Suzanne Carter Kramer, M.Sc.
Thomas A. M. Kramer, M.D.
Mark G. Lebwohl, M.D.
Kenneth Lieberman, M.D.
Charles Lockwood, M.D.
Bernard Mehl, D.P.S.
Myron Miller, M.D.
Edward Raab, M.D.
Allan Reed, M.D.
Allan E. Rubenstein, M.D.
David B. Sachar, M.D.
Henry Sacks, M.D.
Robert Safirstein, M.D.
Ira Sanders, M.D.
Martin H. Savitz, M.D.
Clyde B. Schechter, M.D.
Phyllis Shaw, Ph.D.
George Silvay, M.D.
Barry D. Stimmel, M.D.
Nelson Stone, M.D.
Max Sung, M.D.
Rein Tideiksaar, Ph.D.
Richard P. Wedeen, M.D.
Michael F. Wesson, M.D.

**Editorial Board**

Barry Freedman, M.B.A.
Richard Gorlin, M.D.
Nathan Kase, M.D.
Panayotis G. Katsoyannis, Ph.D.
Charles K. McSherry, M.D.
Jack G. Rabinowitz, M.D.
John W. Rowe, M.D.
Alan L. Schiller, M.D.
Alan L. Silver, M.D.
Alvin S. Teirstein, M.D.

*The Mount Sinai Journal of Medicine* (ISSN No. 0027-2507; USPS 284-860) is published 6 times a year in January, March, May, September, October, and November in one indexed volume by the Committee on Medical Education and Publications (Owner), The Mount Sinai Hospital, Box 1094, One Gustave L. Levy Place, New York, NY 10029-6574. Subscription price: individuals, U.S. and Canada, $60 per year; all libraries, $65; individuals outside the U.S. and Canada, $65. Single copies, $15. New subscriptions begin with the first issue of the calendar year. Send notice of change of address 60 days before the change is effective. Second-class postage paid at New York, NY and at additional mailing offices. POSTMASTER: Send address changes to The Mount Sinai Journal of Medicine, Box 1094, One Gustave L. Levy Place, New York, NY 10029-6574. Copyright 1991 The Mount Sinai Hospital.

*Editorial*

Martin H. Savitz, MD

# Lumbar Disk Disease: Controversies in Neurosurgery

THE ADOPTION of a bipedal erect posture by our hominid ancestors was heralded by an array of effects on the pelvis, hip joints, and spinal column. Five lumbar vertebrae leave the congenitally fused sacrum in a forward direction and then turn sharply back to point upward. Lumbar lordosis carries the torso into a position above the keystone of the pelvis. The hipbones also curve further around the sides toward the front than they do in apes, in order to provide more efficient attachments for the support of the upper trunk. An evolutionary system of ligaments, interposed cartilage, and muscles acts with synergistic and antagonistic precision to hold the vertebrae together and keep the entire column from collapsing. Movement in the normal spine is a smooth and uninterrupted rhythm, free and unrestricted, whether in flexion or extension, from side to side, or rotary.

Low back pain sufficient to require medical attention occurs in 80% of the adult population (1), and 35% of those patients will develop sciatica (2). The incidence of lumbar disk disease in our troglodyte forebears has not been recorded. The image of a cavedweller lying in a flexed position next to the fire to relieve sciatic scoliosis is not so different from a modern patient lying in a hospital bed with heat packs to reduce paravertebral spasm. Homo sapiens has been striding across the savannah for over 50,000 years, but nerve root irritation was not implicated as an etiologic factor in sciatica until the beginning of the twentieth century, and appropriate neurosurgical intervention has only been available for about fifty years.

Sciatica as a symptom complex is attributed to Domenico Cotunnius, a Neopolitan anatomist of the eighteenth century. Lumbago, sciatic neuralgia, and sciatic neuritis were initially differentiated as separate entities. The treatment consisted of varied conservative regimens, all of which included bed rest. With therapy, most patients improved and many recovered completely, while others were subject to recurrent attacks.

Virchow in 1857 described an example of disk protrusion, and Goldthwait in 1911 emphasized the consequences of trauma to the annulus fibrosus resulting in extrusion of the nucleus pulposus (3). Gowers (4) in 1882 reported exostoses of the vertebral bodies compressing the spinal cord. Irritation of the radices spinales was suggested as the cause of sciatic scoliosis by Lortat-Jacob and Sabareanu (5) in 1904 and by Dejerine and Regnard (6) in 1912. Pathologic studies in 1927 by Schmorl (7) showed evidence of prolapse of the nucleus pulposus into the spongiosa of the vertebral body. In 1934, Mixter and Barr (8), and simultaneously the report of Peet and Echols (9), clarified the clinical relationship between herniation of the discus intervertebralis and tension on the nerve root; myelographic study by intraspinal injection of radiopaque oil demonstrated that chondromata were not tumors of cartilaginous origin but prolapses of the nucleus pulposus.

Once rupture of the intervertebral disk had been implicated in impingement on the lumbosacral nerve roots, surgical procedures for removal of the nucleus pulposus became the approach of choice. Formal laminectomy for transdural exploration (10–14) was replaced by laminotomy for diskectomy without the use of the operating microscope (15). Microsurgical diskectomy through a 3-cm incision (16) and minilaminotomy with loupe magnification (17) followed in due course.

Avoiding open operations was advocated in 1960 by Goebert et al. (18), who injected procaine and hydrocortisone into the epidural space of patients with lumbar intervertebral disk syndrome. Hirsch (19) proposed the use of chondrolytic enzymes in the treatment of disk prolapse in 1959, but chemonucleolysis by means of injecting chymopapain into the disk space was not practiced on humans until 1964 by Smith (20). Percutaneous diskectomy by hand was introduced in 1975 by Hijikata et al. (21) and modified to an

---

The senior co-editor of this issue, Professor Leonard I. Malis, has been Neurosurgeon-in-Chief at The Mount Sinai Hospital and Chairman of the Department of Neurosurgery of Mount Sinai School of Medicine since 1970. The junior co-editor of this issue, Dr. Martin H. Savitz, began his neurosurgical residency at The Mount Sinai Hospital in 1970 and is currently Associate Clinical Professor of Neurosurgery at the Mount Sinai School of Medicine and an assistant editor of *The Mount Sinai Journal of Medicine*.

automated procedure by Onik and Maroon (22) in 1988. Arthroscopic diskectomy has been investigated since 1982 by Kambin (23).

Today's neurosurgeon has the benefit of computerized tomographic myelography, magnetic resonance imaging, and an array of other preoperative diagnostic studies. The decision to intervene surgically must be followed by determining which procedure is most suitable for the patient's particular low back dysfunction. Is there one disk herniation producing a single nerve root syndrome that will be relieved by microsurgical excision? Is there extrusion or sequestration of disk material requiring laminotomy, or protrusion that will respond to percutaneous diskectomy? Is decompressive laminectomy and foraminotomy indicated for a stenotic lumbar canal associated with foraminal impingement secondary to facet hypertrophy? If the symptoms are unilateral and the neuroradiologic findings bilateral, should surgery be confined to the side of the pain? If spondylolisthesis or spondylolysis is present at the level of a herniated nucleus pulposus, should spinal fusion be a part of the immediate surgical plan?

After reading the current journal issue and perusing the array of differing opinions, the reader will still be left with a number of quandaries. Each of the authors, after years of experience, remains enthusiastic about the success rate of his particular surgical technique in dealing with a herniated nucleus pulposus. Since practicing neurosurgeons tend to go the way they know—that is, to do the disk procedure most often performed during their training—what is the referring physician to do? A second opinion is unlikely to parrot the advice of the primary surgeon. Correct diagnosis, adequate preoperative neuroradiologic imaging, and skilled relief of nerve root compression are more essential to avoiding failed-back-surgery syndrome than the size of the incision, the amount of bone resected, or the number of grams of disk material removed.

In the span of five decades, all ignorance and controversy about lumbar disk disease has not given way to doctrine. Skepticism toward present entrenched dogma and dissatisfaction with the results of contemporary treatment are sure to beget enough articles to fill the pages of many future journals of medicine and surgery.

## References

1. Nachemson AL. The lumbar spine: an orthopedic challenge. Spine 1976; 1:59–71.
2. Horal J. The clinical appearance of low back disorders in the city of Goetenberg, Sweden. Acta Orthop Scand 1969; [Suppl]118:7–109.
3. Davis L, Davis RA. Principles of neurological surgery. Philadelphia: WB Saunders 1963, 283–291.
4. Gowers WR. Diseases of the nervous system. London: Churchill, 1882, 260.
5. Lortat-Jacob L, Sabareanu C. Sciatique radiculaire inilaterale. Presse Med 1904; 2:633.
6. Dejerine J, Regnard M. Sciatique radiculaire avec paralysie dissociée des muscles antero-externis de la jambe croite: integrite des jambier anterieur: Anesthesie dans le territoire de S.I. Rev Neurol 1912; 23:288.
7. Schmorl C. Ueber die an den Wirbelbandscheiben workommeuden Ausdehnungsund Zerreissungsvorgange und die dadurch an ihnen und der Wirbelspongiosa hervorgerufenen Veranderungen. Verh Dtsch Ges Pathol 1927; 22:250.
8. Mixter WJ, Barr JS. Rupture of the intervertebral disc with involvement of the spinal canal. N Engl J Med 1934; 211:210–215.
9. Peet MM, Echols DH. Herniation of the nucleus pulposus: a cause of compression of the spinal cord. Arch Neurol Psychiatry 1934; 32:92.
10. Elsberg C. Experiences in spinal surgery: observations upon 60 laminectomies for spinal disease. Surg Gynecol Obstet 1913; 16:117–120.
11. Stookey B. Compression of the spinal cord due to ventral extradural cervical chondromas. Arch Neurol Psychiat 1928; 20:275–276.
12. Dandy WE. Loose cartilage from intervertebral disk simulating tumor of the spinal cord. Arch Surg 1929; 19:660–663.
13. Bucy PC. Chondroma of intervertebral disc. JAMA 1930; 94:1552–1553.
14. Finneson BE. Lumbar disc excision. In: Schmidek HH, Sweet WH, eds. Current techniques in operative neurosurgery. New York: Grune and Stratton, 1977, 352–356.
15. Poppen JL. An atlas of neurosurgical techniques. Philadelphia: WB Saunders, 1960, 464–477.
16. Williams RW. Microlumbar discectomy: a conservative surgical approach to the virgin herniated lumbar disc. Spine 1978; 3:175–182.
17. Simeone FA. The neurosurgical approach to lumbar disc disease. Ortho Clin North Am 1971; 2:499–506.
18. Goebert HW, Jallo SJ, Gardner WJ. Sciatica: treatment with epidural injections of procaine and hydrocortisone. Cleve Clin Q 1960; 27:191–197.
19. Hirsch C. Studies on the pathology of low back pain. J Bone Joint Surg 1959; 41B:237–243.
20. Smith L. Enzyme dissolution of the nucleus pulposus in humans. JAMA 1964; 187:137–140.
21. Hijikata S, Yamagishi M, Nakayama T, Oomori K. Percutaneous discectomy: a new treatment method for lumbar disc herniation. J Toden Hosp 1975; 5:5–13.
22. Onik GM, Maroon JC. Percutaneous automated discectomy: a less invasive alternative for the treatment of herniated lumbar discs. Perspect Radiol 1988; 1:1–35.
23. Kambin P, Gellman H. Percutaneous lateral discectomy of the lumbar spine: a preliminary report. Clin Orthop 1983; 174:127–131.

# Conservative Management of Low Back Pain

Richard A. Frieden, M.D.

THE NONOPERATIVE approach to lumbar disk disease begins with the initial contact with the patient who complains of back or leg pain.

## Diagnosis

Distinguishing diskogenic disease from myofascitis or other etiologies is more easily accomplished in the acute stage than in the chronic stage. A careful history and physical examination will usually lead to the diagnosis, or at least to the appropriate tests needed.

Acute back pain due to disk herniation has been described as having sudden onset during flexed lifting, radicular distribution, marked limitation on straight-leg raising, consistent sensory or motor deficits, and depressed or absent reflexes. These characteristics are believed to set disk disease apart from local muscular and skeletal conditions as well as from life-threatening causes—abdominal aortic aneurysms with circulatory collapse, epidural hemorrhage, abscess, or tumor, with sudden onset of paralysis (1).

The severity of disk involvement may be categorized as follows:

- **Mild, giving rise to dorsal ramus syndrome,** pain radiating down the leg to the knee without neurologic findings
- **Major, leading to ventral ramus syndrome,** pain radiating down the leg to the toes, accompanied by neurologic findings on examination.

Classification by symptomatology helps guide further management. However, the presence of a disk bulge may be an incidental finding on an imaging study and may not correlate with the patient's complaints and the clinician's findings on examination.

The questionnaire on pain is an attempt to characterize the nature and effect of the patient's pain. The patient is encouraged to categorize the discomfort rather than adopt an anxious, helpless approach, which can impede treatment. The drawing of pain sites on the questionnaire helps establish the diagnosis, since precise boundaries suggest radicular involvement, and diffuse markings suggest myofascial problems.

When viewing the patient's posture and movements, the clinician must bear in mind the anatomy that allows or restricts motion (2). Much information can be obtained without ordering imaging studies by testing for

- physiologic responses
- dynamic vs. static activity
- systemic diseases

No one test will determine the best approach to treatment. The treating physician must consider the whole picture.

**Testing Physiologic Responses.** Thermography has been reported to delineate myotomes and dermatomes affected by compression of the nerve root. This rather controversial technique cannot distinguish disk disease from other types of spinal pathology. A negative thermogram, however, may suggest that further testing is unnecessary because no pathophysiologic process is present.

Electrodiagnostic studies (electromyography, nerve conduction velocities, somatosensory evoked potentials) can demonstrate the presence of nerve injury, level of injury, degree of damage, and whether the process is acute or chronic in nature. The results of tests have been used to justify the performance of many surgical procedures. However, these techniques are fraught with potentially confounding factors. Although positive sharp waves and fibrillation potentials are defi-

---

From the Department of Rehabilitation Medicine, Mount Sinai School of Medicine. Address reprint requests to the author at The Mount Sinai Medical Center, One Gustave L. Levy Place, Box 1240, New York, NY 10029.

nite signs of abnormality, they can be misinterpreted if confused with low amplitude, short-duration motor unit discharges.

The paraspinal musculature is the first to show evidence of radiculopathy, and it is the first group to "normalize" as chronicity develops. Myelography, epidural injections, and nerve root blocks, however, can elicit spontaneous activity. In addition, the paraspinal musculature can only be used to determine the level of nerve root damage if the examination includes the multifidus muscle. If the level is not verified at the time of examination, the examiner must rely on the distal (limb) musculature to determine the level of involvement (3).

The examination yield is limited by the expertise of the examiner. The method is nondiagnostic in the postlaminectomy patient. Anomalies of nerve origin and of muscle innervation may confuse the picture. Although electrodiagnostic studies can identify radiculopathy, they may not identify or localize the cause.

**Testing Dynamic vs. Static Responses.** The full nature of spinal mobility and its role in back pain and disk disease remains unknown. Investigators have claimed that characteristic patterns of muscle use and spinal motion correlate with different causes of back pain (4–7). Increased interest in the measurement of vertebral motion has led to the development of machines that avoid the radiation of motion fluoroscopy.

The IsoStation B-200 is one example that assesses three axes of range of motion by reporting position, velocity, and torque (8).

An alternative to a restrictive frame is the Spinoscope, which uses a complement of light-emitting diodes and surface electromyographic electrodes to measure range of motion and muscle activity while allowing greater freedom of motion (9).

On a simpler scale, observation of the patient's posture and movements behind a transparent grid can help quantify these measurements (2).

**Testing for Systemic Diseases.** The patient's history itself may suggest a possible systemic cause of back pain. Various blood tests may be helpful, including serologic tests for autoimmune diseases, antigen typing if HLA-B27 is suspected, and tumor markers.

## Treatment

The proposed modes of conservative treatment of lumbar disk disease are as varied as the proposed mechanisms for its pathogenesis.

**Bed Rest and Traction.** The use of bed rest as a treatment modality during the acute stage is beneficial because it allows the inflammatory reaction to resolve. Complete rest also relieves stress on the paraspinal muscles. The herniated disk may not physically recede, but intradisk pressure is lowest in the supine position.

The ultimate effect of bed rest depends on many factors:

- Patients who are allowed to decide for themselves when to end the period of bed rest have poorer outcomes than do patients with a predetermined period of bed rest.
- Length of time for bed rest; recommendations have ranged from two days to two months. Prolonged bed rest has significant deleterious effects on many body systems, including the musculoskeletal system. Deyo et al. (10) suggested that only two days of bed rest may be needed for acute mechanical back pain, followed by structured resumption of work. For the patient with disk herniation and neurologic deficits, a longer period may be warranted. Fast (11) in 1988 agreed that bed rest was needed for persons with disk herniation and signs of neuropathy.
- Positioning while in bed. A variety of rolls, wedges, and pillows are available as aids.

Because compliance with orders for "strict bed rest" is usually poor at home, many patients are admitted to hospital for bed rest and pelvic traction. For actual vertebral space widening, the weight used for pelvic traction would have to exceed 40%–50% of body weight; otherwise, traction is simply a means to enforce bed rest. Gravity lumbar reduction is a form of head-up axial traction that works by fixating the rib cage and allowing the rest of the body (approximately 40% of body weight) to hang down (12). A combination of suspension by the hands from a bar-rack and intermittent head-down traction on a tilt-board for postoperative patients and as part of conservative management was reported by Potthoff et al. (13).

**Injections.** Either dry needling without injection of medications, or local infiltration of anesthetic agents, can provide relatively quick, sometimes lasting, relief of pain when carefully defined trigger points are found. Steroids may also be helpful; repeated injections are often necessary for optimal effect.

Epidural steroid-anesthetic injections have been advocated for many years for their anti-inflammatory and analgesic effects.

**External Support.** Bracing of the trunk is usually not indicated for low back pain. There is an inverse correlation between the degree of support and the degree of comfort. Reliance on a thoracolumbar spinal orthosis may lead to further weakening of trunk and back muscles and an increased risk of re-injury. The patient who wears an orthosis must be encouraged to continue exercises.

**Physical Modalities.** Mobilization techniques include mechanical manual traction, vertebral segment joint mobilization, and numerous forms of massage and soft-tissue manipulation. Other modalities include superficial heat and cold, deep heating with ultrasound, and muscle stimulation by electric currents.

Transcutaneous electrical nerve stimulation (TENS) has a place in the management of many types of acute and chronic pain. TENS is approximately 60% to 80% effective. However, a recent article by Deyo et al. (14) suggested that treatment with TENS is no more effective than a placebo for patients with chronic lumbar back pain.

**Exercises.** For patients who have chronic nonspecific back pain or subacute subsiding episodes, general fitness should be encouraged. Daily routines of general conditioning may begin with superficial heat and TENS if needed. The patient should start with exercises in the supine position and then advance to the standing position.

The choice of strict flexion exercises (Williams) versus starting with extension and gradually including other positions (McKenzie) has been a source of controversy for many years (15). Although Williams flexion exercises can strengthen abdominal muscles, stretch back muscles and soft tissues, and open intervertebral foramina, they can raise intradisk pressure.

McKenzie exercises consist primarily of extension exercises; the patient is encouraged to seek positions that reduce pain and centralize it, eliminating pain from the legs (11).

**Education: The "Back School."** After the acute pain has begun to subside, the next step in the rehabilitation of the patient with lumbar disk disease is education—posture analysis, problem-solving at home and work to anticipate potential reinjury situations, information on the anatomy of the back and proper lifting techniques, and an exercise program. Patients who understand the mechanisms of back motion and injury may have better long-term outcomes as well as more motivation to take an active role in rehabilitation.

The "back school" was originally organized as a series of lectures and demonstrations that covered anatomy, physiology, posture, relaxation, lifting, carrying, and the importance of regular exercise. The primary goals are to educate patients and to encourage them to take an active role in the treatment and prevention of back disorders (16), through several components:

- screening for appropriate patients
- education
- monitoring effectiveness
- motivating patients to emphasize self-care over work skills

The importance of patient motivation cannot be overemphasized. Suffering results from the cultural perception of pain. Behavior that reduces pain can be reinforced by a compensation system or reduced by a comprehensive program of pain management (17). A program of behavioral modification helps the patient distinguish pain from the underlying physiologic processes (18).

The "back school" begins in the emergency room. When the patient has gone home and the acute phase has subsided, the "pupil" is then ready to learn preventive measures. After the patient has begun a course of rehabilitation, work hardening may be introduced. A "return-to-work" program uses work simulation and ergonomics to retrain patients.

Patients may also be managed in a general pain clinic with a comprehensive multidisciplinary approach.

**Length of Trial of Conservative Management.** If the initial physical examination and testing do not reveal an emergent or surgically correctable problem, conservative measures should be attempted for six months unless neurologic deterioration occurs.

## Efficacy of Nonoperative Care

In 1989, Saal and Saal studied patients with documented lumbar disk herniations without significant stenosis who underwent an aggressive program including back school, general conditioning, and stabilization exercises. The results indicated that herniated disks can be treated successfully without surgery if the patient takes an active stance toward treatment. The suggestion was also made that failure to improve with aggressive nonoperative treatment indicated the presence of spinal stenosis (19).

Rosomoff (20) prescribed intensive physical therapy preoperatively for his patients with documented spinal stenosis; the multidisciplinary approach resulted in less surgery.

The first mode of treatment of back or leg pain, whether acute or long term, should be aggressive rehabilitation.

### References

1. Carron H. Acute low back pain. Emergency Decisions 1986; 2(7):26–39.
2. Russek AS. Biomechanical and physiological basis for ambulatory treatment of low back pain. Orthop Rev 1976; 5(4):21–31.
3. Saal JA, Saal JS. Electrophysiologic evaluation of lumbar pain: establishing the rationale for therapeutic management. In: White AH, Rothman RH, Ray CD, eds. Lumbar spine surgery: techniques and complications. St. Louis: CV Mosby, 1987.
4. Smith SS, Mayer TG, Gatchel RJ, et al. Quantification of lumbar function. Part 1: isometric and multispeed isokinetic trunk strength measures in sagittal and axial planes in normal subjects. Spine 1985; 10:757–764.
5. Mayer TG, Smith SS, Keeley J, et al. Quantification of lumbar function. Part 2: sagittal plane trunk strength in chronic low-back patients. Spine 1985; 10:765–772.
6. Mayer TG, Smith SS, Kondraske G, et al. Quantification of lumbar function. Part 3: preliminary data on isokinetic torso rotation testing with myoelectric spectral analysis in normal and low-back pain patients. Spine 1985; 10:912–920.
7. Kishino ND, Mayer TG, Gatchel RJ, et al. Quantification of lumbar function. Part 4: isometric and isokinetic lifting simulation in normal subjects and low-back dysfunction patients. Spine 1985; 10:921–927.
8. Isotechnologies, Inc. POB 1239, Hillsborough, NC 27278.
9. Spinex Medical Technologies, Inc. 1980 Sherbrooke ouest, Montreal, Canada H3H 1E8.
10. Deyo RA, Diehl AK, Rosenthal M. How many days of bed rest for low back pain? A randomized clinical trial. N Engl J Med 1986; 315:1064–1070.
11. Fast A. Low back disorders: conservative management. Arch Phys Med Rehabil 1988; 69:880–891.
12. Burton CV, Kirkaldy-Willis WH, Yong-hing K, Heithoff KB. Causes of failure of surgery on the lumbar spine. Clin Orthop Rel Res 1981; 157:191–199.
13. Potthoff PC, Greitemann B. Early results of systemic extension therapy in patients with herniated lumbar discs and similar conditions. In: Wullenweber R, Linger M, Brock M, eds. Advances in neurosurgery. Berlin: Springer, 1987.
14. Deyo RA, Walsh NE, Martin DC, et al. A controlled trial of transcutaneous electric nerve stimulation (TENS) and exercise for chronic low back pain. N Engl J Med 1990; 322:1627–1634.
15. Ponte DJ, Jensen GJ, Kent BE. A preliminary report on the use of the McKenzie protocol versus Williams protocol in the treatment of low back pain. J Orthop Sports Phys Ther 1984; 6:130–139.
16. Forssell MZ. The back school. Spine 1981; 6:104–106.
17. Fordyce WE. Pain and suffering—a reappraisal. Am Psychol 1988; 43:276–282.
18. Fordyce WE, Brockway J, Bergman JA, Spengler D. A control group comparison of behavioral vs. traditional management in acute back pain. J Behav Med 1986; 5(2):127–140.
19. Saal JA, Saal JS. Nonoperative treatment of herniated lumbar intervertebral disc with radiculopathy: an outcome study. Spine 1989; 14(4):431–437.
20. Rosomoff HL, Rosomoff RS. Nonsurgical aggressive treatment of lumbar spinal stenosis. SPINE: State of the Art Reviews 1987; 1(3):383–399.

# Management of Low Back Pain by Analgesics and Adjuvant Drugs

JOSEPH R. DIPALMA, M.D., AND G. JOHN DIGREGORIO, M.D., PH.D.

MORE THAN other fields of therapy, the management of low back pain with analgesic drugs and their adjuvants must rely on art as much as on science, for several reasons. For one, the pain is disabling; for another, it is persistent and chronic. But the main reason for the need to rely on art is that a specific etiology (with a specific remedy) is identified in only 50% of cases (1, 2). Low back pain is often associated with an overlay of anxiety, depression, and feelings of inferiority (3, 4).

The perfect analgesic drug that controls pain without adverse reactions and addictive properties has yet to be devised (5). It should be obvious that drug therapy for pain must fit into and complement the main therapeutic approach for a particular patient. For example, does one prescribe an analgesic with sedative properties when exercise is the prime therapeutic approach? Are opiates a good choice when the nature of the disorder is going to be chronic and recurrent? When and why are anti-inflammatory drugs indicated?

## Drug Selection

In a broad sense, management of drugs for low back pain is no different from management of pain in other disorders. The approach and rationale must be based on symptoms and cause when these are known. Although an algorithm could be constructed for specific patients, the complexity and diversity of factors in low back pain make more practical a classification based on duration of symptoms, accurate assessment of pain symptoms, and response to treatment (Table 1) (6).

The type of patient who has acute symptoms is usually a young to middle-aged adult with an initial attack of low back pain which may have followed heavy exercise or may simply have arisen spontaneously. Even without therapy, one can expect recovery in a week or two. Symptoms recur only at long intervals—months and years—and there is no obvious associated abnormality. Such patients will require only nonnarcotic analgesics such as acetaminophen, or nonsteroidal antiinflammatory agents (NSAIDs). Occasionally the pain is so severe that no rest can be obtained in any position, nor can exercise be performed, in which case a narcotic such as codeine, oxycodone, or meperidine may be added. The narcotic should be continued for not more than 10 days. In addition, it is important to ration the narcotic to the degree of pain. Gradually reducing the dose as the pain subsides is wise. Rarely are muscle relaxants or other adjuvant drugs used.

Postoperative pain from corrective spinal surgery is another category of acute low back pain symptoms. Here, especially, in the few days immediately after the operation no hesitation should be exercised in giving morphine and other powerful opiates such as methadone, by the parenteral route initially and orally later. Addiction is rarely a problem, provided the usual precautions are observed. NSAIDs are also administered and are often helpful and sparing of the dosage of opiates. Only in exceptional circumstances are muscle relaxants and psychotropic agents useful in treating postoperative pain.

Pain caused by specific neural involvement generally responds poorly or only partially to analgesics. In any event, analgesics do not remedy the loss of motor and reflex function. However, some such patients do recover with medical measures without surgery; some may refuse surgery.

---

From the Department of Pharmacology, Hahnemann University School of Medicine, Philadelphia, PA. Address reprint requests to G. John DiGregorio, M.D., PhD, Hahnemann University School of Medicine, M.S. #431, Broad and Vine Streets, Philadelphia, PA 19102.

TABLE 1

*Selection of Drugs by Symptomatology to Treat Low Back Pain*

| Nature of pain symptoms | Drugs of choice |
|---|---|
| Acute | |
| Short duration | NSAIDS |
| Infrequent | Acetaminophen |
| No obvious associated pathology | Opiates, usually milder (Codeine, oxycodone) |
| Diagnosis frequently obscure | Rarely adjuvant drugs |
| Postoperative pain | |
| Chronic | |
| Frequent or persistent | NSAIDS (plus GI protective agents) |
| Associated pathology (eg, generalized arthritis) | Opiates (rarely; only for short periods) |
| Spine disorders, disk hernia spondylolysis, spondylolisthesis, lumbar facet syndrome, segmental instability, spinal stenosis, disk degeneration | Adjuvant drugs: muscle relaxants, antidepressants, antianxiety and neuroleptics, rarely corticosteroids |
| Intractable | |
| Failed-laminectomy syndrome | Epidural morphine and corticosteroids |
| Neoplastic disease | Adjuvant drugs, phenothiazines, antihistamines |
| Neurologic disorders | Patient self-assisted analgesia |
| Psychological factors | Referral to pain clinic |

Here drug therapy aims to correct the associated pain-reinforcement factors so often present in chronic pain syndromes.

**Nonsteroidal Anti-Inflammatory Drugs.** Local pain is generated by an inflammatory process that includes hyperemia, edema, the attraction and concentration of leukocytes and lymphocytes, and the secretion of local control substances (autocoids), including bradykinin, prostaglandins, serotonin, lymphokines, and many other peptides, such as the central nervous system mediator substance P. Of these, the first three and substance P have definitely been proven to cause pain (7, 8). Prostaglandin synthesis inhibition by aspirin and NSAIDs has proven to be a most effective method of analgesia for pain produced secondary to a local inflammatory process (9). Apparently the phenomenon of response to local injury, whether mechanical, infectious, chemical, or autoimmune, is the same and always induces prostaglandins. Actually the prostaglandins themselves do not directly produce pain; they act by inducing a state of hyperalgesia caused by activation of bradykinins. Hence, NSAIDs have proven to be universally useful to control the inflammatory process and thus contribute considerable and useful analgesic action, no matter what the etiology of the disorder (Table 2).

**Opiates.** Opiates may be combined with NSAIDs to achieve a degree of analgesia exceeding that of NSAIDs used alone. The main adverse actions of opiates are respiratory depression, constipation, and addiction. An opiate in combination with an NSAID lowers the analgesic dose of each component, which is a decided advantage. Thus codeine, oxycodone, and propoxyphene are available combined with aspirin or acetaminophen in convenient dosage forms. These are popular and widely prescribed (Percocet, Percodan). However, the opiate content is no less addicting when combined with NSAIDs, and such preparations must be used cautiously and with limitations and for acute temporary situations (Tables 3, 4).

**Skeletal Muscle Relaxants.** Muscle relaxants have a limited role in low back pain (Table 5). Muscle spasm most often occurs with injury and may be a reflex caused by the pain itself. Tenderness and pain on palpation, along with increased muscle tone, must be present. Spasm should be distinguished from spasticity caused by central nervous system diseases such as multiple sclerosis and cerebral palsy. As a rule direct-acting or spinal-level muscle relaxants are not useful in these disorders.

Muscle spasm is a spinal reflex and ideally skeletal muscle relaxing drugs should inhibit the internuncial neurons at a spinal segmental level and have little or no effect on higher levels of the central nervous system. Unfortunately, the available skeletal muscle relaxants, such as carisprodol, chlorzoxazone, and chlorphenesin, cause drowsiness, sedation, light-headedness, and dizziness. In addition, tolerance develops quickly; large doses taken over weeks or months lead to physical and psychological dependence resembling barbiturate dependency (5).

The skeletal muscle relaxants are most useful combined with bed rest and physical therapy in acute situations such as trauma where spasm is obviously present and therapy is to last only for a week or two. This therapeutic approach is promoted by drug manufacturers and many preparations of skeletal muscle relaxants combined with aspirin or acetaminophen are available (10). A recent double-blind study with tizandidine (a new skeletal muscle relaxant) combined with ibuprofen in acute low-back pain showed only slight advantage over ibuprofen alone (11). Fixed drug combinations are convenient in general practice but do not allow the clinician to use acumen and sagacity to the maximum advantage of the patient.

**Antidepressants.** The use of antidepressant drugs as adjuvant analgesics has gained wide acceptance (12–14). An original concept was that depression reinforced the pain reflex and that relief of this mood would aid analgesic drugs in the

TABLE 2

*Nonsteroidal Anti-inflammatory Drugs (NSAIDs) Available for Low Back Pain*

| Drug | Trade names | Dose, mg/day | | Analgesic potency | Form Rx | Comment |
|---|---|---|---|---|---|---|
| | | Starting | Maximum | | | |
| Salicylates | | | | | | |
| Acetyl salicylic acid | Aspirin (Various) | 1400 | 6000 | + | Buffered enteric coated with antacids | Standard of NSAID therapy; for most cases as effective as newer drugs |
| Choline magnesium, salicylate | Trilisate | 1500 × 1 then 1000 | 4000 | + | Tablets (500, 750, 1000 mg)<br>Liquid | May have less GI toxicity than other NSAIDs. Magnesium may be toxic in renal insufficiency. As with all salicylates, danger of "salicylism" |
| Magnesium salicylate | (Various) | 2600 | 4800 | + | Tablets (325, 500, 545, 600 mg) | |
| Diflunisal | Dolobid | 1000 | 1500 | + | Tablets (250, 500 mg) | May have less GI toxicity than aspirin |
| Salsalate | (Various) | 1500 × 1 then 1000 | 4000 | + | Tablets (500, 750 mg) | Insoluble in acid secretions, hence not absorbed till it reaches intestine |
| Proprionic acids | | | | | | |
| Ibuprofen | Motrin, Advil (Various) | 1200 | 3200 | + | Tabs 200, 300, 400 mg | Available over the counter in 200 mg tablets |
| Naproxen | Anaprox, Naprosyn | 500 | 1000 | + | Tabs 250, 275 mg | |
| Naproxen sodium | Anaprox DS | 550 | 1500 | + | Tab 550 mg | |
| Fenoprofen | Nalfon (Various) | 800 | 3200 | + | 200, 300, 600 mg capsules | |
| Acetic acids | | | | | | |
| Indomethacin | Indocin | 75 | 200 | + + | 25, 50, 75 mg<br>Caps suspension<br>Suppositories | More potent analgesic; more toxic to GI tract and CNS |
| Tolmetin | Tolectin | 600 | 2000 | + | 200, 400, 600 mg Tabs | |
| Sulindac | Clinoril | 300 | 400 | + | 150, 200 mg Tabs | Claims less renal toxicity |
| Diclofenac | Voltaren | 100 | 200 | + | Tabs 25, 50, 75 mg | |
| Oxicams | | | | | | |
| Piroxicam | Feldene | 20 | 40 | + + | Caps 10 mg | More potent analgesia; more toxic |
| Fenamates | | | | | | |
| Mefamamic acid | Ponstel | 500 | 1000 | + | Caps 250 mg | Generally not recommended for long-term use |
| Meclofenamic acid | (Various) | 150 | 400 | + | Caps 50, 100 mg | |
| Pyrazoles | | | | | | |
| Phenylbutazone | Butazolidin (Various) | 300 | 400 | + + + | Tabs & caps 100 mg | Older drug; the most potent analgesic & anti-inflammatory; toxicity causing aplastic anemia and agranulocytosis precludes general use |
| p-Amino-phenol derivative | | | | | | |
| Acetaminophen | Tylenol (Various) | 1400 | 6000 | + | Tabs 80, 120, 160, 325 mg<br>Caps 500 mg<br>Elixir<br>Suppositories | Does not affect platelet aggregation; no GI irritation; lacks antiinflammatory action; does not inhibit prostaglandin synthesis |

abatement of pain, especially chronic pain. Early attempts in this direction included the use of caffeine, dextroamphetamine, and even cocaine in combination with aspirin. However, these CNS stimulants are not true antidepressants.

Drugs which inhibit the uptake of monoamines (serotonin, norepinephrine) in the CNS, such as tricyclics (TCAs), are legitimate antidepressants (5). In their own right TCAs can inhibit endogenous monoamine-dependent pain modulating systems. Clearly recognized is the serotonergic tract traveling from the medulla to the dorsal horn of the spinal cord, involving an enkephalinergic internuncial neurone which functions to block nociceptive signals from the periphery (15). With regard to this pathway the most

TABLE 3
*Commonly Prescribed Opiate Agonists*

| Drug | Trade names | Equianalgesic doses, mg | | t½ (hr) | Duration of action (oral) | Comparative effects | | |
|---|---|---|---|---|---|---|---|---|
| | | IM | Oral | | | Analgesic | Respiratory depression | Physical dependence |
| Morphine | (Various) | 10 | 60 | 2–4 | 3–7 | + + | + + | + + |
| Levorphanol | Levo-Dromoran | 2 | 4 | 12–16 | 4–8 | + + | + + | + + |
| Hydromorphone | Dilaudid | 1.5 | 7.5 | 2–3 | 4–5 | + + | + + | + + |
| Oxymorphone | Numorphan | 1 | 6 | — | 3–6 | + + | + + + | + + + |
| Methadone | Dolophine | 10 | 20 | 15–30 | 4–6 | + + | + + | + + |
| Meperidine | Demerol<br>(Various) | 75 | 100 | 3–4 | 2–4 | + + | + + | + + |
| Fentanyl | Sublimaze<br>(Various) | 0.1 | — | 1.5–6 | 1–2 | + | + | — |
| Codeine | (Various) | 130 | 200 | 3 | 4–6 | + | + | + |
| Oxycodone | Roxicodone | — | 30 | — | 4–6 | + + | + + | + + |
| Propoxyphene | Darvon<br>(Various) | — | 130 | 6 to 12 | 4–6 | + | + | + |

effective adjuvant analgesic TCAs are tertiary amine compounds such as amitriptyline, which have greater activity on serotonin uptake than on norepinephrine uptake. However, the analgesic mechanism is probably more complex, because other pain pathways appear to be inhibited. Pain relief occurs before mood changes in patients with chronic pain receiving TCAs (Table 6).

Thus TCAs, although not a panacea, are worth a trial in the patient who has chronic pain, especially but not only if the patient has depression. The preferred drugs are amitriptyline, doxepin, and imipramine, but other TCAs may also be effective.

**Miscellaneous Adjuvants.** When low back pain includes a component of neuropathic pain characterized as lancinating, certain anticonvulsants have been found to be beneficial (16). The theory underlying anticonvulsive therapy for this type of pain is that these agents control nociceptive pulses by a variety of mechanisms, neuronal hyperexcitability, suppression of paroxysmal discharges, and spread of abnormal depolarization (17). Such drug therapy has gained acceptance for trigeminal neuralgia using phenytoin or carbamazepine. More recently, other anticonvulsant drugs have been successfully used to treat other syndromes of lancinating pain, such as diabetic neuropathy, tabetic lightning pains, multiple sclerosis paroxysmal pain, postsympathectomy pain, cancer pain, and posttraumatic neuralgia (18).

Lancinating pains are uncommon in the ordinary low back pain syndrome but do occur in neoplastic disease of the spine, multiple sclerosis, and posttraumatic and postsurgical pain syndromes. When the lancinating pain is chronic and is causing much distress, adding anticonvulsive drugs to the regimen of pain control appears worthwhile. Recently, oral antiarrhythmic drugs such as tocainide and mexiletine have also been shown to relieve lancinating pain (19, 20). Naturally, experience with these agents for this use is limited.

In controlled trials, antihistamines have shown analgesic effects (21). The trials were of diphenhydramine and hydroxyzine, but there is no reason to believe that other antihistamines would not have the same analgesic effect. Hydroxyzine, which has marked sedative as well as other psychic actions, has received special study (22). Despite much promotion, the antihistamines have not gained much acceptance as adjuvants for pain relief. For cancer pain associated with anxiety and nausea, hydroxyzine is at times useful as an adjuvant to standard analgesics (23).

Theoretically the neuroleptic drugs, pheno-

TABLE 4
*Opioid Agonist-Antagonists*

| Drug | Trade name | Analgesic dose equivalent to 10 mg IM Morphine (mg) | Relative antagonist activity | Mode of administration | Duration of action (hr) |
|---|---|---|---|---|---|
| Pentazocine | Talwin | 30–60 | 1 (weak) | IM, IV, oral | 3 |
| Butorphanol | Stadol | 2 | 10–30 | IM | 3–4 |
| Nalbuphine | Nubain | 10 | 10 | IM, IV | 3–6 |
| Buprenorphine | Buprenex | 0.3 | very strong, equipotent with naloxone | IM, IV | 6 |

thiazine and butyrophenones, for example, should be useful adjuncts to analgesic drugs. In general their CNS depressive actions are additive to opiates; the neuroleptics are also antiemetic, which makes them useful in preanesthetic medication (5). However, neuroleptics as analgesics have for the most part only anecdotal evidentiary support (24). They have been used in various types of neuropathy, thalamic pain, and arthritis.

One phenothiazine, methotrimeprazine, has been developed as an analgesic. As a potent CNS depressant it produces suppression of sensory impulses, reduction of motor activity, sedation, and tranquilization. It also raises the pain threshold, accompanied by amnesia to a degree equivalent to morphine (10, 25). Methotrimeprazine is available only as an intravenous preparation and its marked sedative effects preclude its use outside hospitals.

Thus, with the exception of methotrimeprazine, the neuroleptic drugs have only a small role to play as analgesic adjunctive drugs. Their use is restricted to patients who have the most refractory and severely incapacitating low back pain, usually the result of advanced neoplastic disease.

**Agents for Refractory, Intractable Pain.** There are a small fraction of patients whose low back symptoms become intractable and completely incapacitating (26). They usually have a history of surgery which has failed and they can name all the drugs that have not worked. Opiate addiction is a frequent complication. Various names are applied to this group, such as "failed laminectomy syndrome" and "low-back losers" (27).

Corticosteroids by various routes constitute one of the drug therapies of last resort for patients with intractable pain. In most cases systemic corticosteroids are considered to be too toxic. Depot forms of the steroid plus a local anesthetic are injected into the suspected spinal area. The original concept was to reduce local inflammation, but there is some evidence that locally applied adrenal steroids produce a profound and long-lasting block of neuroma hyperexcitability (28). Methylprednisolone acetate, which has low solubility and hence a prolonged effect, may be injected in a dose of up to 80 mg into the soft tissue or intraarticularly at the site of a mechanosensitive "trigger point."

Where there is arthritic involvement in low back pain, some authorities feel systemic corticosteroid therapy is indicated. The results are always good, perhaps because the corticosteroids induce a feeling of well-being along with the lessening of inflammation. The problem is that corticosteroid therapy is always toxic (leading to gastric ulceration and bleeding and Cushing's syndrome) and it is very difficult to withdraw the patient from the therapy (29). If such therapy is to be used, the danger must be explained to the patient. Pulse dosage or alternate-day dosage is claimed to be less toxic and more manageable (30). Considering these difficulties, systemic corticosteroids are to be used only in the most extreme cases.

TABLE 5

*Centrally Acting Skeletal-Muscle Relaxants*

| Drug | Trade name | Usual oral dose total daily (mg) | Dose interval (hr) |
|---|---|---|---|
| Carisoprodol | Soma<br>(Various) | 1400 | QID<br>4 |
| Chlorphenerin carbamate | Maolate | 1600 | QID<br>4 |
| Chlorzoxazone | Paraflex<br>(Various) | 1000 | QID<br>4 |
| Cyclobenzaprine | Flexeril | 40 | QID<br>4 |
| Metaxalone | Skelaxin | 3200 | QID<br>4 |
| Methocarbamol | Robaxin<br>(Various) | Initial 6000<br>Maintenance 4000 | QID<br>4 |
| Orphenadrine | Norflex<br>(Various) | 200 | BID<br>6 |
| Diazepam | Valium<br>(Various) | 20 | QID<br>4 |

At the juncture when all the analgesic and adjuvant drugs have failed and invasive and mechanical and electrical manipulations have also failed, the physician will refer the patient to a pain clinic. Here the advantage of integrated team work of psychologists, physicians, skilled nurses and physical therapists can concentrate on a more exact diagnosis of the cause of the pain and hence more accurate and specific therapy. Some patients can be helped by more skillful use of systemic drug therapy, but many patients will end up with epidural morphine alone, or with a corticosteroid, usually methylprednisolone (31). This therapy is claimed to produce long-term (6–24 months) pain relief in many patients. As expected, there is no change in preexisting abnormalities, and there are skeptics (32).

Epidural morphine has an established history of successful use as an analgesic (33, 34). There is evidence that morphine has a direct effect on the spinal cord in blocking pain transmission (35). There are, of course, enkephalic (opiate) receptors in the spinal cord. The advantage of epidermal versus systemic parenteral administration of morphine is that a suitable dose may be given locally with minimal serious side effects such as respiratory depression. The mystery is the long-term effect when combined with corticosteroids, an effect less enthusiastic observers do

TABLE 6
*Antidepressants Useful as Adjuvants for the Relief of Pain*

| Drug | Trade names | Usual oral dose total daily (mg) | Dose interval (hr) | Activity | | |
|---|---|---|---|---|---|---|
| | | | | Serotonin-uptake blocking | Anticholinergic | Sedation |
| Amitriptyline | Elavil, Amitril<br>(Various) | 75 | 4 | + + + + | + + + + | + + + |
| Impramine | Trofamil<br>(Various) | 75 | 4 | + + + + | + + | + + |
| Doxepin | Adapin, Sinequan<br>(Various) | 150 | 4 | + + | + + | + + + |

not seem able to duplicate (32). A more recent innovation is "patient-assisted analgesia"; the patient controls the amount of epidural morphine injected with the aid of an electromechanical device.

Many other drugs have received some attention for the relief of low back pain that does not respond to common analgesics. A recent example is colchicine, the well-established alkaloid useful in the therapy of acute gouty arthritis (36). Several double-blind studies have shown it to be highly effective when administered intravenously but not orally (37). The mechanism of action in disk diseases is uncertain but has been attributed to inhibition of the formation of leukotrine B (38).

## General Considerations

In this report the concentration has been on the indications for drug use, and clinical pharmacologic information has been limited because of limitations of space. The skillful use of drugs requires this information; fortunately, a number of handy pocket-sized manuals provide it (39, 40). Reference to the package insert or other comprehensive manuals is preferred when convenient (16). A number of excellent short works on pain assessment and management (41–43) are also available.

Some helpful suggestions which apply to each drug group follow.

**Nonsteroidal Anti-Inflammatory Drugs.** Long-term use of NSAIDs in large doses usually results in gastric irritation, ulceration, and sometimes serious bleeding. Attempts to counteract these effects have led to buffered, enteric coated and compounded tablets containing antacids. Other gastric mucosal protectants are sucralfate and the $H_2$ blockers, such as cimetidine and ranitidine. The most recent antidote for this serious problem is the prostaglandin $E_1$ analog, misoprostol, approved by the Food and Drug Administration for the prevention of NSAID-induced gastric ulcers in high-risk patients (44). In clinical studies, misoprostol is at least as effective as $H_2$ blockers in healing idiopathic gastric ulcers, but it has greater side effects because a higher dose must be used. Other than diarrhea, the main problem with misoprostol is its high cost, limiting its use to the wealthiest patients (45).

Among the NSAIDs, at least 17 compounds are effective (Table 2). Which to select for a particular patient depends on many factors. For patients with average uncomplicated conditions, any salicylate would be an obvious choice; a good substitute is one of the proprionic agents. Sulindac, which has less renal toxicity than other NSAIDs, may be preferred for patients with renal damage. When a higher degree of analgesia is required, piroxicam or indomethacin might be indicated, but they are more toxic. Phenylbutazone, an older drug, is the most potent analgesic, but should rarely be used because of its severe toxicity to bone marrow. Acetaminophen (not an NSAID) is a good substitute for mild analgesia but has no anti-inflammatory action. It has the great advantage of lacking the gastrointestinal toxicity of the NSAIDs.

**Opiates.** The clinician who has learned to use morphine expertly has little use for the many synthetic analogs (Table 3). Morphine is poorly absorbed by the oral route, and the does must be at least 6 times the intramuscular dose. Methadone is relatively well absorbed orally and has a long half-life. This makes it an ideal oral analgesic drug for cancer patients who must take opiates for long periods. Meperidine also has the disadvantage of a short half-life because it stimulates its own metabolism. The addiction potential is related directly to the potency of the opiate. Weaker analgesics such as meperidine and oxycodone have undeserved popularity, since at equipotent analgesic action they have comparable adverse reactions of respiratory depression and physical dependence. The opiates that are more potent weight for weight, hydromorphone, oxymorphone, and fentanyl, have specialized uses and are rarely resorted to for general analgesia.

The opioid agonist-antagonists (Table 4) were developed with the goal of providing good analgesia with little or no development of tolerance and hence of physical dependence. Much research has isolated selective actions of agonist and antagonist drugs on opiate receptors mu, kappa, and sigma. The hope was that combining agonist and antagonistic properties in the new synthetic opiate compounds could achieve analgesia without a potential for abuse. The results have fallen short of the mark. Nevertheless, the agonist-antagonists have achieved definite utility in patients who have already shown dependence susceptibility. Only pentazocine is available for oral use. Mainly these agents are used in acute in-hospital situations. All of them can precipitate an alarming withdrawal syndrome in patients who have previously been taking an opioid agonist. Caution is to be exercised: agonist-antagonist drugs should not be administered to individuals who have been on chronic opioid therapy. Acute alarming withdrawal symptoms will occur unless the agonist opioid has been completely eliminated.

The practical problem with opiate analgesia is the management of chronic pain from nonmalignant causes. Although society and the medical profession accept the use of addicting opiate therapy for someone who is certainly going to die of his or her disease, they frown on its use to treat persons in pain who are going to live for years.

Opioid maintenance therapy for pain arising from nonmalignant disorders has gained more acceptability in the past 5 years (46–48). The problem is that such therapy places an extraordinary responsibility on the physician and the patient. In maintenance therapy, opiates are suitable for only a few carefully selected patients. Guidelines, ethical and professional, for such selection have been stated (42). A more suitable solution might be the assignment of the task of opiate chronic pain management to an institutional review board or an ethics committee. In carefully selected cases, long-term management of nonmalignant pain via opiates should be considered as an alternate to invasive and destructive procedures that often have little promise of success for intractable pain.

## Summary

Ninety percent of all low back pain can be managed successfully by the skillful and appropriate use of the major analgesic drugs, NSAIDs, and the opioids. Adjunctive drugs, such as skeletal muscle relaxants and antidepressants, have their uses, but must be considered as temporary aids to the major analgesics.

There is intensive research to develop new and more useful analgesic drugs. Ketorolac tromethamine, an NSAID recently approved by the Food and Drug Administration, is suitable for intramuscular administration for the short-term management of postsurgical and other acute pain (49).

## References

1. Waddell G. A new clinical model for the treatment of low-back pain. Spine 1987; 12:632–644.
2. Quebec Task Force on Spinal Disorders. Scientific approach to the assessment and management of activity-related disorders: a monograph for clinicians. Spine 1987; 12(Suppl.):522–530.
3. Waddell G, Morris EW, DiPaola M, Bircher M, Finlayson D. A concept of illness tested as an improved basis for surgical decisions in low-back disorders. Spine 1986; 11:712–719.
4. Mechanic D. Illness behaviour, social adaptation and the management of illness. J Neuro Ment Dis 1977; 165:79–87.
5. DiPalma JR, DiGregorio GJ, eds. Basic pharmacology in medicine, 3rd ed. New York: McGraw-Hill, 1990, 293–323.
6. Lee CK. Office management of low back pain. Ortho Clin North Am 188; 19:797–804.
7. Cohen DH, Sherman SM. The somatosensory system. In: Berne RM, Levy MN, eds. Physiology, med ed. St. Louis: Mosby, 1988, 146–147.
8. Alberto B, Brag D, Lewis J, Raff M, Roberts K, Watson JD. Molecular biology of the cell, 2nd ed. New York: Garland, 1989, 689, 974.
9. Cooper SA. New peripherally-acting oral analgesic agents. Am Rev Pharmacol Toxicol 1983; 23:617–647.
10. Drug facts and comparisons: skeletal muscle relaxant combinations. St. Louis: JB Lippincott, 1990, 1328.
11. Berry H, Hutchinson DR. Tizanidine and ibuprofen in acute low back pain: results of a double-blind multicentre study in general practice. J Int Med Res 1988; 16:83–91.
12. Walsh TD. Antidepressants in chronic pain. Clin Neuropharmacol 1983; 6:271–295.
13. Butler S. Present status of tricyclic antidepressants in chronic pain. In: Benedetti C, Chapman CR, Mosicca G, eds. Advances in pain research and therapy, vol. 7 New York: Raven Press, 1984, 173–198.
14. Getto CJ, Soskness CA, Howell T. Antidepressants and chronic nonmalignant pain: a review. J Pain Manage 1987; 2:91–98.
15. Hammond DL. Pharmacology of central pain-modulating networks (bigenic amines and nonopioid analgesics). In: Fields HL, Dubner R, Cervero F, eds. Advances in pain research and therapy, vol. 9. New York: Raven Press, 1985, 499–513.
16. Swerdlow M. Anticonvulsant drugs and chronic pain. Clin Neuropharmacol 1984; 7:51–82.
17. Weinberger J, Nicklas WJ, Berl S. Mechanism of action of anticonvulsants. Neurology 1976; 26:162–173.
18. Swerdlow M, Cundil JG. Anticonvulsant drugs used in the treatment of lancinating pains: a comparison. Anesthesia 1981; 36:1129–1132.

19. Lindstrom P, Lindblom U. The analgesia effect of tocainide in trigeminal neuralgia. Pain 1987; 28:45–50.
20. Dejgard A, Petersen P, Kastrup J. Mexiletine for treatment of chronic painful diabetic neuropathy. Lancet 1988; 1:9–11.
21. Rumore MM, Schlichting DA. Clinical efficacy of antihistamines as analgesics. Pain 1986; 25:7–22.
22. Bellville JW, Dorey E, Capparell D, et al. Analgesic effects of hydroxyzine compared to morphine in man. J Clin Pharmacol Therapeutics 1979; 19:290–296.
23. Foley KM. The treatment of cancer pain. N Engl J Med 1985; 84–95.
24. Kocher R. Use of psychotropic drugs for the treatment of chronic pain. In: Bonica JJ et al., eds. Advances in pain research and therapy, vol. 1. New York: Raven Press, 1976, 279–282.
25. Beaver WT, Wallenstein SM, Honde RW, Rogers AA. A comparison of the analgesic effects of methotrimeprazine and morphine in patients with cancer. Clin Pharmacol Therapeutics 1966; 7:436–446.
26. Auld AW, Murdoch DM, O'Laughlin KA. Intrapinal narcotic analgesia; pain management in the failed laminectomy syndrome. Intraspinal Narcotic Analgesia 1987; 12:953–954.
27. Belkin S. Back pain in evaluation and treatment of chronic pain. In: Aronoff GM, ed. Baltimore: Urban and Schwarzenberg, 1985, 343–350.
28. Devor M, Govrin-Lippmann R, Raber P. Corticosteroids suppress ectopic neuronal discharge originating in experimental neuromas. Pain 1985; 22:127–137.
29. Kountz DS. An algorithm for corticosteroid withdrawal. Am Fam Physician 1989; 39:250–254.
30. Zarro VJ. Steroid pulse therapy. Am Fam Physician 1986; 33:217–219.
31. Cohn ML, Huntington CT, Byrd SE, et al. Epidural morphine and methylprednisolone, new therapy for recurrent low-back pain. Spine 1986; 11:960–963.
32. Brechner VL. Letter to the Editor. Spine 1987; 12:827.
33. Behar M, Magora F, Olshwang D, Davidson DT. Epidural morphine in treatment of pain. Lancet 1979; 1:527–529.
34. Bromage PR, Camporisi E, Chestnut D. Epidural narcotics for postoperative analgesia. Anesth Analg 1980; 59:473–480.
35. Yoksh TL, Rudy TA. Analgesia mediated by a direct spinal action of narcotics. Science 1976; 192:1357–1358.
36. Meek JB, Guidice VW, Enrick NL. Colchicine highly effective in disk disorders. J Neurolog Orthopaed Surg 1984; 5:215–220.
37. Schnebel BE, Simmons JW. The use of oral colchicine for low-back pain: a double blind study. Spine 1988; 13:354–357.
38. Francis AE. Theoretical considerations for the use of colchicine (articulorum) in disk diseases. J Neurolog Orthopaed Surg 1986; 7:235–251.
39. DiGregorio GJ, Barbieri GJ, Sterling GH, Camp JF. Handbook of pain management, 2nd ed. Westchester, Pa: Medical Surveillance, 1989.
40. Aronoff GM, Evans WO. Handbook on the rational use of medication for pain. New York: Lawrence A. Della-Corte Publications, 1987.
41. Fields HL, ed. Pain syndromes in neurology. Boston: Butterworths, 1990, 257, 278.
42. Carnic PM, Brown FD, eds. Assessing chronic pain: a multidisciplinary handbook. New York: Springer-Verlag, 1988.
43. Ghose K, ed. Drug management of pain in the elderly. Boston, MA: MTP Press, 1987.
44. DiPalma JR. Misoprostol: a prostaglandin for peptic ulcer. Am Fam Physician 1989; 40:217–219.
45. Edelson JT, Tosteson ANA, Sax P. Cost-effectiveness of misoprostol for prophylaxis against nonsteroidal antiinflammatory drug-induced gastrointestinal tract bleeding. JAMA 1990; 264:41–53.
46. Protenoy RK, Foley KM. Chronic use of opioid analgesics in nonmalignant pain: report of 38 cases. Pain 1986; 25:171–186.
47. France RD, Urban BJ, Keefe FJ. Long-term use of narcotic analgesics in chronic pain. Soc Sci Med 1984; 19:1379–1382.
48. Taub A. Opioid analgesics in the treatment of chronic intractable pain of non-neoplastic origin. In: Kitahata LM, Collins D, eds. Narcotic analgesics in anesthesiology Baltimore: Williams and Wilkins, 1982, 199–208.
49. O'Hara DA, Fragen RJ, Kinzer M, Pemberton D. Ketorolac tromethamine as compared with morphine for the treatment of postoperative pain. Clin Pharmacol Ther 1987; 41:556–61.

# Arthritis of the Lumbar Spine

GARY V. GORDON, M.D., F.A.C.P.

LOW back pain falls under the purview of internists as well as neurosurgeons and orthopedists. In Great Britain, back pain accounts for 4.2% of all general-practice consultations (1). In many rheumatologists' practices, 40% of the patients consult the physician for low back pain. Prior to the introduction of magnetic resonance imaging, at least $16 billion was spent in the United States annually on back pain (2). These facts no doubt inspired Sir William Osler's comment, "When a patient comes in the front door with arthritis, I go out the back door."

Although pervasive and frustrating to treat, back pain is usually a benign condition. Nonetheless, treatment may be impeded by a thick veneer of psychological and legal factors—including lost wages, claims for worker's compensation, physicians' and lawyers' fees, and hospital and medication costs.

Like the blind man and the elephant, the physician is never quite sure what part of the conundrum he or she is holding. There is often no way to determine whether low back pain is due to degenerative joint disease, apophyseal joint synovitis, muscle spasm or strain, thickened ligamentum flavum, or any of the other possibilities (3). With the exception of the nucleus pulposus, the spinal disks are innervated structures and thus potential sources of pain (4). This review focuses on the contribution of arthritis to low back pain from the viewpoint of a rheumatologist.

## Differential Diagnosis

Before focusing on the treatment of the syndrome of arthritic low back pain, it is important to review the mental gymnastics involved in differential diagnosis. Possibilities include mechanical, inflammatory, visceral, infectious, neoplastic, and metabolic causes (see ruled box).

A history of corticosteroid use may suggest osteoporosis and compression fractures. Night pain and morning stiffness in a young man are a clue to ankylosing spondylitis, and an HLA-B27 blood test may be helpful in diagnosis. Fever should alert the physician to consider endocarditis or osteomyelitis. The bone pain of metastases or myeloma tends to be more continuous, even during recumbent periods (5). A rectal examination is useful to rule out prostatic metastases. Alkaline and acid phosphatase, sedimentation rate, calcium, and hemoglobin levels may shed light on systemic disorders.

As a practical point, plain view radiographs are generally too frequently ordered and are rarely cost effective (6). The documentation of osteophytes or disk degeneration rarely alters the clinical course. However, back pain correlates more closely with disk space narrowing than with osteophytes (7). Bone destruction from tumor or infection is extremely rare but is the major observation that can profoundly change therapy.

Most back pain is mechanical and, fortunately, short-lived. Our ability to detect arthritic changes radiographically is not matched by similar clinical acumen. Although much low back pain remains undiagnosed (8), fortunately 90% of patients have pain resolution within two months (9). However, the 5% who still have low back pain after three months account for 85% of the cost in terms of work time lost and compensation benefits (10). These patients often come to our offices after seeing several physicians, physical therapists, and chiropractors. Many of these patients have degenerative disorders, but 50% may have no structural diagnosis. Half of this undiagnosed group may score high in anxiety, hysteria, and hypochondriasis on the Minnesota Multiphasic

---

From the Department of Medicine, Thomas Jefferson University College of Medicine, Philadelphia, PA. Address reprint requests to Gary V. Gordon, M.D., 320 Lankenau Medical Building, Wynnewood, PA 19096.

**Classification of Low Back Pain**

**Mechanical**
Strain/sprain
Degenerative disk disease
Disk herniation
Spinal stenosis
Fracture

**Infectious**
Vertebral osteomyelitis
Endocarditis

**Inflammatory**
Ankylosing spondylitis
Reiter's disease

**Metabolic**
Osteoporosis
Paget's disease

**Visceral**
Pancreatic cancer
Pyelonephritis
Aortic aneurysm

**Neoplastic**
Myeloma
Metastatic malignancy

Personality Inventory (11), and in this group therapeutic triumphs are likely to be infrequent.

Arthritis in low back pain manifests itself primarily as a degenerative disorder, osteoarthritis. An overlapping disease, diffuse idiopathic skeletal hyperostosis, and ankylosing spondylosis account for a small percentage of low back pain. Osteoarthritis involves both the intervertebral disk space and the apophyseal joint (spondylosis). Extensive apophyseal involvement can ultimately cause spinal stenosis by compromising the central spinal canal, lateral recesses, and the neural foramina.

**Pathology: Intervertebral Osteoarthritis.** Disk degeneration becomes noticeable by age 20. The nucleus pulposus is soft and gelatinous at birth, but eventually is replaced by fibrocartilage. The cartilaginous end-plate, which is closely applied to the central depressions of vertebral bodies, eventually fibrillates and fissures without repair. The anulus is initially attached to the anterior and posterior longitudinal ligaments, but loses its lamellar shape by the fifth decade and fissures. Schmorl and Junghans argue that breakdown of the anular fibers allows disk material to displace outward, resulting in protrusion (12) (Fig. 1). The process is accentuated when the nucleus pulposus is normal and saturated with fluid. As the disk desiccates with time, there is less pressure within the nucleus pulposus and protruding disks cause fewer symptoms in the elderly population. Neuropathic symptoms in these patients stem more from apophyseal disease and spinal stenosis when new bone growth impinges on nerve roots.

With disruption of anular fibers, the nucleus pulposus prolapses into the epidural space. Most rheumatologists view bulging of the disk as a radiographic curiosity in search of clinical expression. When disks herniate, pieces may break off as sequestered or free fragments. Even when the pathology is discreet, the mechanism of root pain is still not clear, since root compression is not always found at surgery. Inflammation of involved nerve roots, which occurs universally and has been substantiated by biopsy (13), most likely contributes to the perception of pain.

Other potential contributors to low back pain are related to the death of matrix cells in a degenerating disk. Enzymes are released that can destroy collagen fibers and cause mechanical tension on irritated nerve endings (14). In addition, as disk proteoglycans degrade, breakdown products may activate Hageman factor and eventually release kinins (15) that stimulate pain in free nerve endings.

An intriguing hypothesis implicating venous obstruction in the intervertebral foramen as the cause of nerve root pain was recently suggested by Hoyland et al. (16). These investigators dissected 160 lumbar foramens in cadavers and found 73% to have intraforaminal venous dilatation secondary to disk or osteophyte compression and direct nerve compression in only 8 specimens. Although venous compression and neurofibrosis correlated, medical records in this retrospective study did not allow good clinical correlation.

**Biochemical Changes.** Disk cartilage degeneration is a primary factor in spinal arthritis. Cartilage is composed principally of water, proteoglycans, and collagen. All three components are known to change with age: Water is progressively lost, collagen cross-linking deteriorates, and the smooth lamellar lines of collagen on the hyaline cartilage surface become disrupted. Inappropriate production of type I collagen can be found in degenerating nuclei (17). Proteoglycans, normally long elastic proteins, may be reduced in size and chemically altered with more keratin sulfate side chains. Nutritional factors may play a role, as increased levels of lactate are found with the onset of anaerobic metabolism. Degenerating disks, particularly in the elderly, may calcify with mineral deposits in the anulus, which becomes more brittle and decreases spine flexibility.

## Pathology: Apophyseal Joints

Some classicists have contended that osteoarthritis of the spine refers only to the apophyseal joint because its synovium is similar to that of peripheral joints. With cartilage fibrillation and erosion, the synovium in the apophyseal joint becomes secondarily irritated (18). Autopsy studies have demonstrated the frequency of synovial inflammation and hypertrophy in osteoarthritis (19). The revisionist view contends that the inflamed synovium plays a primary role in releasing proteolytic enzymes destructive to cartilage (20).

Current opinion favors an interaction between cartilage and synovium leading to further joint degeneration. Shards of cartilage embedded in synovium, as well as calcium hydroxyapatite and calcium pyrophosphate, have been identified as dihydrate inciting an inflammatory response (19). Cooke (21) has suggested that immune complexes may be trapped in the upper zone of articular cartilage and may perpetuate the inflammatory response.

Synovitis has a twofold significance. Cartilage breakdown is a source of potentially reversible pain. Apophyseal synovitis is responsive to nonsteroidal anti-inflammatory drugs. Consequently, some patients with low back pain may improve as this inflammation is controlled.

Osteoarthritis is a dynamic process. As cartilage deteriorates, new bone forms, perhaps as a poor attempt to compensate for the loss of joint surface. When bone spurs occur in peripheral joints (for example, Heberden nodes), they seldom create significant problems other than loss of full joint range of motion. In the spine, however, bone spurs compromise foramina of exiting nerve roots as well as the spinal canal. As facet joints hypertrophy, the protruding new bone invaginates into the spinal canal and distorts this tube from an oval to a trefoil shape where the lateral recesses are pinched (Fig. 2). Corresponding clinical symptoms of neurogenic claudication (with normal peripheral pulses) and positional back pain with multilevel electromyographic abnormalities may be expected. In such cases, nonsteroidal anti-inflammatory drugs and other pharmacologic interventions have little to offer.

Biomechanically, disk degeneration and apophyseal disease may be viewed as interacting in a continuum. As disk cartilage deteriorates, loss of disk height develops. With disk space shrinkage, there is posterior displacement of the vertebral body and subsequent subluxation of the apophyseal joint. Apophyseal joints take over as a fulcrum without the normal sliding motion, a change that can lead to slipping, listhesis, and deterioration (22).

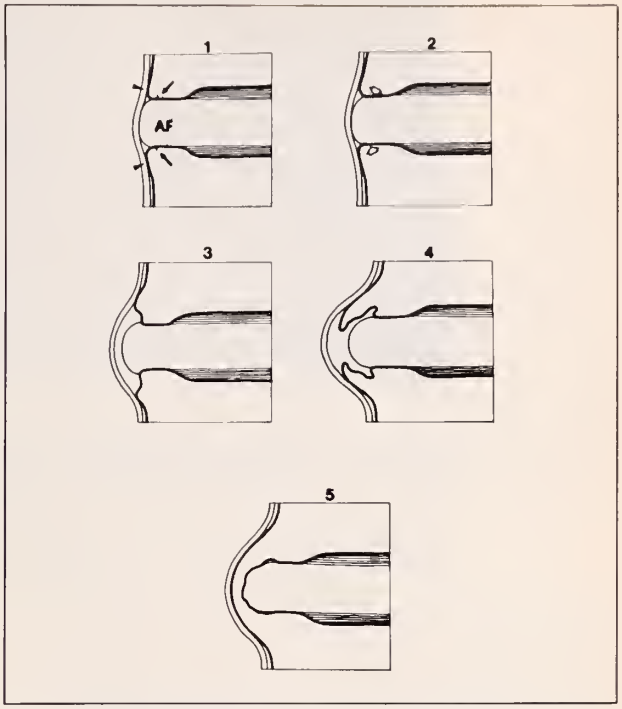


**Fig. 1.** Progressive stages of spondylosis deformans: normal, drawing 1, upper left; sequential abnormalities, drawings 2–5. Note normal attachment of anulus fibrosus (AF) to vertebral rim by strong fibers (solid arrows) and attachment of anterior longitudinal ligament to anterior edge of vertebral bodies (arrowheads). With breakdown in site of attachment of anulus fibrosus to vertebral body (open arrows), anterior protrusion of intervertebral disk material produces elevation of anterior longitudinal ligament and enlarging osteophytes at point of ligamentous attachment to vertebral body (2–4). Eventually, osteophytes may bridge intervertebral disk space (5). *Reproduced* by permission from Resnick and Niwayama (ref. 18).

## Spondyloarthropathies

Although osteoarthritis accounts for the majority of joint disorders causing low back pain, other forms of arthritis may contribute, particularly in men in their third and fourth decades. The spondyloarthropathies, for lack of a better term, encompass such diseases as ankylosing spondylitis, Reiter's disease, and psoriatic arthritis. Many of these patients carry the genetic histocompatibility antigen HLA-B27; the gene occurs in about 6%–7% of the white population and in 3% of blacks. There is approximately 20% penetrance in men and far lower expression in women. In theory, 1.5 million men in the United States may have symptomatic ankylosing spondylitis. Young men typically have night pain and morning stiffness, whereas osteoarthritis symptoms tend to get worse with weight bearing. Ra-

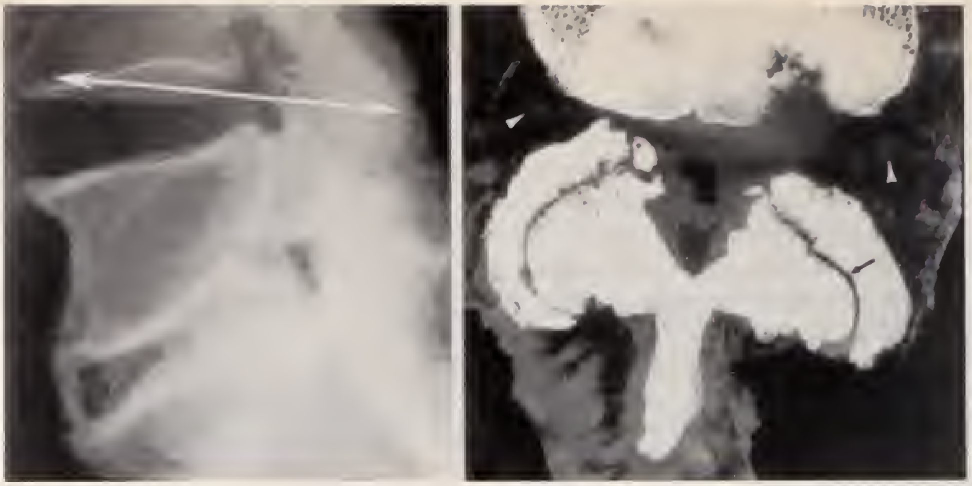


**Fig. 2.** Apophyseal joint osteoarthritis: lumbar spine. **(Left)** Lateral radiograph of cadaveric spine, showing considerable alteration of apophyseal articulations between fourth and fifth lumbar vertebrae and between fifth lumbar vertebra and sacrum. Joint space narrowing, bone sclerosis, and osteophytes are apparent. **(Right)** Radiograph of axial section at approximate level indicated by double-headed arrow (left) reveals osteoarthritic changes. Findings include joint space narrowing, vacuum phenomenon (arrow). Although neural foramina are narrowed, there is no impingement on nerve roots (arrowheads), which have passed through foramina at higher level. *Reproduced* by permission from Resnick and Niwayama (ref 18).

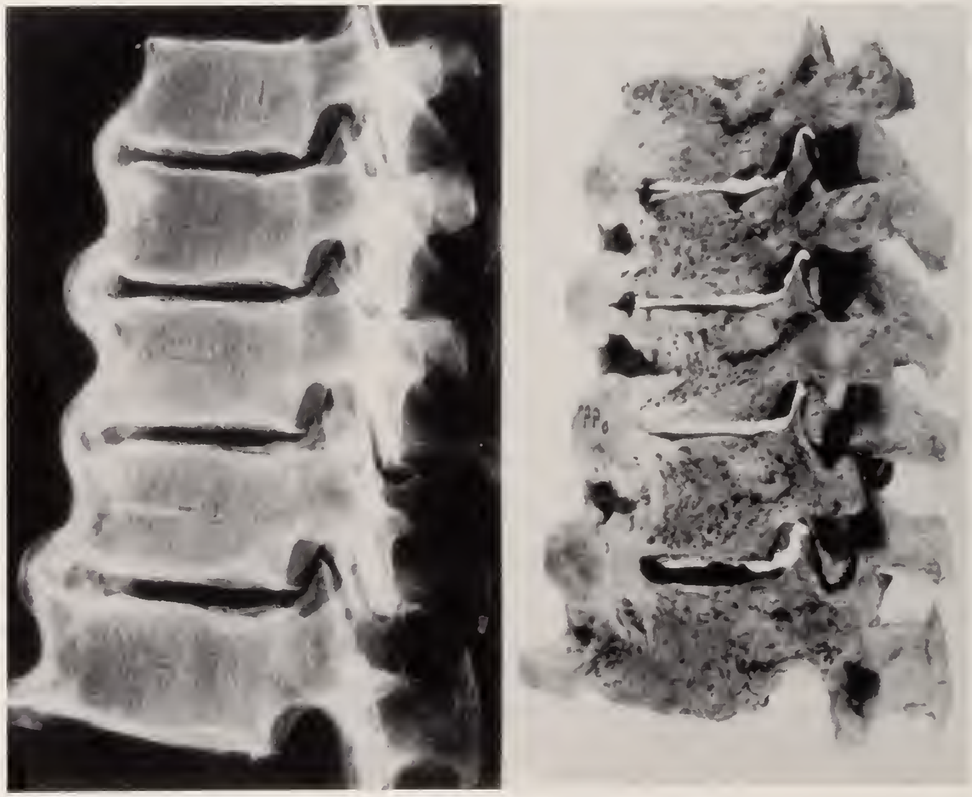


**Fig. 3.** Diffuse idiopathic skeletal hyperostosis. Radiograph **(left)** and photograph **(right)** of lower thoracic spine and thoracolumbar junction; anterior ossification has produced undulating vertebral contour. Osseous excrescences are largest at levels of intervertebral disks and midportions of vertebral bodies are less affected. Hyperostosis of one spinous process is seen. *Reproduced* by permission from Resnick and Niwayama (ref. 18).

diographs may be normal at the onset, leading some patients to be misdiagnosed as malingerers. Early lesions begin in the sacroiliac joint, extend up the spine, and sometimes involve C1–2. Bone scanning may detect early changes, but normal increased uptake in incompletely mature sacroiliac joints may lead to incorrect evaluation.

The primary target of the spondyloarthropathies is not cartilage, as in osteoarthritis, but rather the entheses, where spinal longitudinal ligaments insert. These insertion sites occur along the vertebrae and are normally fibrous. The enthesitis of ankylosing spondylitis involves an inflammatory reaction followed by calcification and, ultimately, ossification of the longitudinal ligaments. The vertebrae become squared off as the superior and inferior vertebral margins are eroded; the process results in the classic “bamboo spine.”

Treatment includes use of anti-inflammatory medications, as well as exercises to prevent deformities such as thoracic kyphosis. Other diseases in this family, such as psoriatic arthritis and Reiter’s disease, have less spinal involvement. Asymmetric involvement of the lumbar spine occurs in these disorders, whereas the ossified vertical ligaments (syndesmophytes) of ankylosing spondylitis are bilateral. Synovitis plays a larger role in Reiter’s and psoriatic arthritis and may be amenable to drugs other than the NSAIDs (methotrexate, for example).

### Diffuse Idiopathic Skeletal Hyperostosis

Diffuse idiopathic skeletal hyperostosis is a disease of new bone formation in the thoracolumbar spine. Large, beaked, horizontal spurs occur at many levels, particularly on the right side (Fig. 3). The disk space remains intact, but the spurs may ankylose the spine and cause stiffness. This entity has features mildly suggestive of ankylosing spondylitis, although the spurs are more typical of an exhuberant form of osteoarthritis. HLA-B27 testing will be positive in 34% of these patients. The posterior elements (vertebral arch and apophyseal joint) of the spinal column are spared, and flexibility is minimally compromised. The spurs are not clinically significant except in the cervical region, where impingement on the esophagus may cause dysphagia.

“Whiskering” about the pelvis can be seen in this disease, and peripheral joints such as the calcaneus and olecranon may have large bone spurs. Large hyperostoses may be painful in joints or in tendinous or ligamentous attachments.

### Summary

Many types of arthritis can disturb the normal lumbar anatomy. This review intentionally omits many endocrine, metabolic, and genetic disorders that ultimately lead to breakdown of cartilage. Osteoarthritis remains the most common cause of sustained back pain as well as one of the most compensable disabilities in the United States. As so many other authors have noted, more scientific investigation is needed, and funding is imperative to effect significant changes.

### References

1. Wood PNH. Epidemiology of back pain. In: Jayson M, ed. The lumbar spine and back pain. London: Sector, 1976.
2. Grazier KL, Holbrook TL, Kelsey JL, Stauffer RN, eds. The frequency of occurrence, impact and cost of musculoskeletal conditions in the United States. Chicago: American Academy of Orthopedic Surgeons, 1984.
3. Nachemson A. Towards a better understanding of low back pain: a review of the mechanics of the lumbar disc. Rheum Rehab 1975; 14:129–143.
4. Hirsch C, Ingelmar RBE, Miller M. The anatomical basis for low back pain. Acta Ortho Scanda 1963; 33:1–17.
5. Hadler NM. Regional back pain. N Engl J Med 1986; 315:1090–1092.
6. Deyo RA. Plain roentgenography for low back pain. Arch Int Med 1985; 149:27–28.
7. Togerson WR, Dotter WE. Comparative roentgenographic study of the asymptomatic and symptomatic lumbar spine. J Bone Joint Surg 1976; 58:50–55.
8. Nachemson A. The lumbar spine—an orthopedic challenge. Spine 1976; 1:59–71.
9. Dillane JB, Fry J, Kalton G. Acute back syndrome—a study from general practice. Br Med J 1966; 3:82–84.
10. Frymoyer JW, Pope MH, Clements JH, et al. Risk factors in low back pain: an epidemiological survey. J Bone Joint Surg 1983; 65A:213–218.
11. Pope MH, Rosen JC, Wilder DG, Frymoyer JW. The relation between biochemical and psychological factors in patients with low back pain. Spine 1980; 5:173–178.
12. Schmorl G, Junghans H. The human spine in health and disease. 2nd ed. Translated by EF Besemaum. New York: Grune & Stratton, 1971:138.
13. Murphy RW. Nerve roots and spinal nerves in degenerative disc disease. Clin Orthop 1977; 129–147.
14. Brown MD. The source of low back pain and sciatica. Sem Arthritis Rheum 1989; 18(4, Suppl 2):67–72.
15. Moskowitz RW, Schwartz HJ, Michael B, et al. Generation of kinin-like agents by chondroitin sulfate, heparin, chitin sulfate, and human articular cartilage: possible pathophysiologic implications. J Lab Clin Med 1970; 76:790–798.
16. Hoyland JA, Freemont AJ, Jayson MIV. Intervertebral foramen venous obstruction: a cause of periradicular fibrosis? Spine 1989; 14:558–568.

17. Nimni ME. Collagen: structure, function and metabolism in normal and fibrotic tissues. Sem Arthritis Rheum 1983; 8(1):1–86.
18. Resnick D, Niwayama G. Diagnosis of bone and joint disorders, 2nd ed. Philadelphia: WB Saunders, 1988.
19. Gordon GV, Villanueva T, Schumacher HR, Gohel V. Autopsy study correlating degree of osteoarthritis, synovitis, and evidence of articular calcification. J Rheum 1984; 11:681–686.
20. Glynn LE. Primary lesion in osteoarthritis. Lancet 1977; 1:574–575.
21. Cooke TD. Significance of immune complex deposits in osteoarthritic cartilage. J Rheum 1987; (Spec No):77–79.
22. Farfan HF, Cossette JW, Robertson GH, et al. Effects of torsion on the lumbar intervertebral joints: the role of torsion on the production of disc degeneration. J Bone Joint Surg 1970; 52:468–475.

# Low Back Pain in Children

Bruce R. Rosenblum, M.D., and Allen S. Rothman, M.D.

## Abstract

Low back pain is a rare phenomenon in children. The etiology differs from the degenerative and diskogenic disease in adults and is almost never a manifestation of psychological disturbance. Low back pain in children is often the harbinger of serious underlying pathologic changes and should be evaluated by state-of-the-art neuroradiologic tests.

Sciatica, although a frequent complaint among the adult population, rarely occurs in children. Low back pain is the most common cause of hospitalization among adults, accounts for one-quarter of lost work time, and is second only to respiratory infections in causing sick days (1, 2).

Although only one-tenth of all low back pain can be traced to specific causes (3), the majority of cases in adults with a known cause appear to be related to muscular strain, degenerative disease, and diskogenic disorders. A large proportion of these dysfunctional states, both acute and chronic, are secondary to work-related injuries (3).

In contradistinction, the maladies that affect children and subsequently manifest themselves as low back pain are distinctly different. Degenerative changes are rare in a child's intervertebral disk and spinal column (4). Children and adolescents account for less than 2% of the reported cases of lumbar disk herniation (4). This fact has been attributed to the progressive dehydration of the intervertebral disk with advancing age (88% water content in childhood, compared to 69% hydration by the eighth decade). Lumbar disk degeneration first appears on pathologic examination in its mildest form in 11–19-year-old males and a decade later in females (5, 6); recent confirmation has been reported in studies utilizing magnetic resonance imaging (MRI) proton density images to measure water content (7, 8).

The complaint of low back pain in young children must therefore be regarded as highly significant and possibly indicative of serious underlying disorder with minimal clinical signs and apparently unrelated symptoms (9, 10). A frequently incorrect assumption is that low back pain in children is a psychogenic disorder related to either malingering or secondary gain (4). Remembering that children rarely manifest emotional problems as low back pain can avoid delayed or missed diagnoses (10).

## Evaluation

Although history-taking remains an important element in the evaluation of the child with low back pain, the patient's responses may frequently be less than satisfying. Children are poorly able to localize the site of their pain along the spinal axis. More commonly, they simply report discomfort "in the back." The pain often seems to be referred to the lower back despite the location of the pathologic process in more cephalad portions of the spinal column. The phenomenon may be due, in part, to the creation of subclinical or overt misalignment of the developing spine by a distant process. Although initially due to asymmetric function of the paravertebral musculature in an attempt to "splint" the spine and diminish pain, a fixed deformity may develop if the underlying process is left untreated.

Physical examination of the child with low back pain differs from that of the adult in several

---

From the Department of Neurosurgery, The Mount Sinai School of Medicine, New York, NY. Address reprint requests to Allen S. Rothman, M.D., 1160 Fifth Avenue, Suite 106, New York, NY 10029.

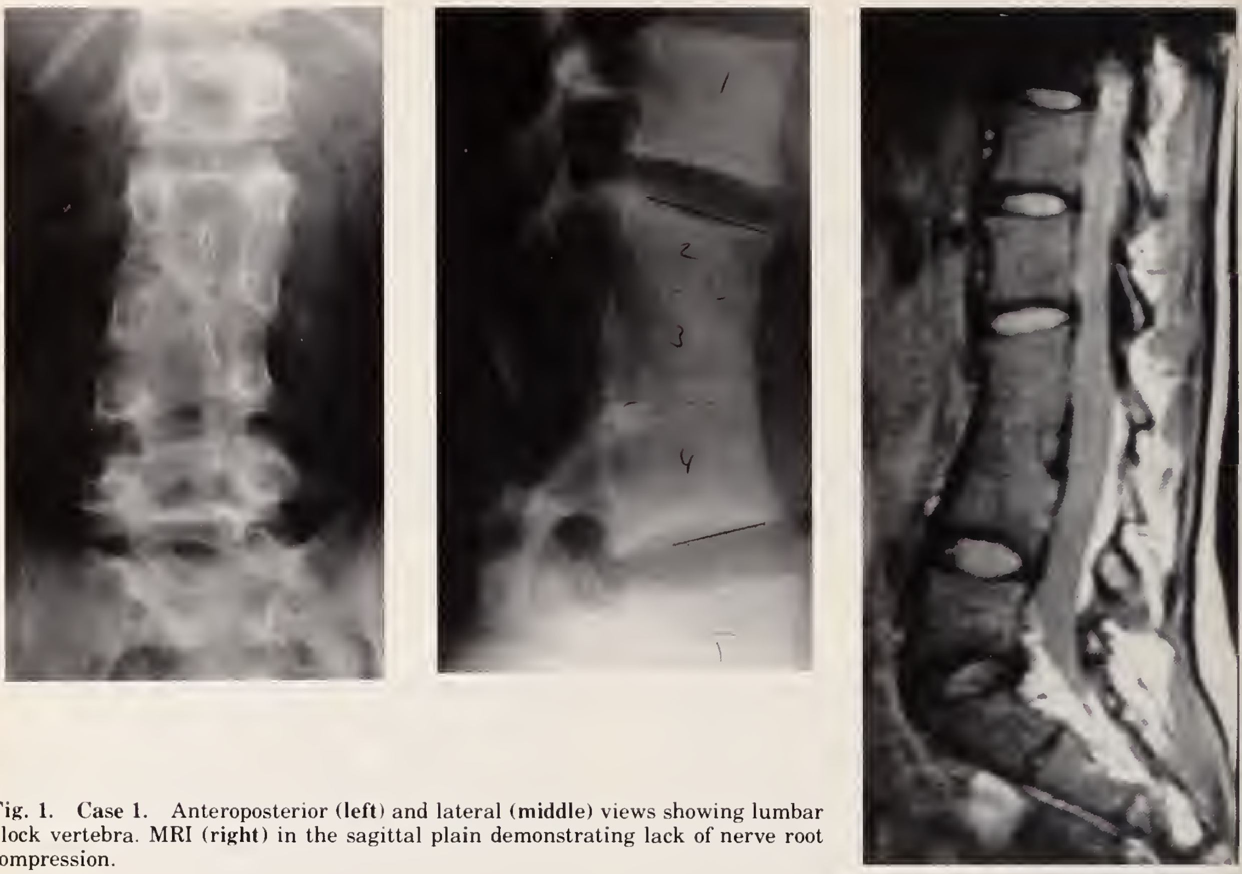


Fig. 1. Case 1. Anteroposterior (**left**) and lateral (**middle**) views showing lumbar block vertebra. MRI (**right**) in the sagittal plain demonstrating lack of nerve root compression.

important respects. Posture should be assessed, as should the presence and extent of spinal-column deformities. Although scoliosis or other spinal curvatures may be idiopathic, particularly if present from birth, there may also be a clue to a pathologic condition involving the contents of the spinal canal. Neuroradiologic studies should be performed in any child harboring an abnormal curvature of the spinal column.

The skin, particularly overlying the spinal axis, should be inspected for stigmata of neurocutaneous syndromes—fistulae, dimples, subcutaneous masses, tufts of hair, or abnormal pigmentation. These signs may be harbingers of spinal neoplastic or vascular disorders or dysraphic states. Mobility of the lumbosacral spine should be tested for restriction of movement, a sign of spinal cord tethering and related congenital malformations.

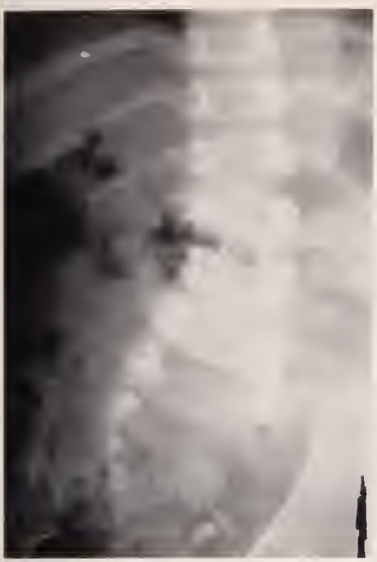
Neuroradiologic investigation of low back pain in children has been revolutionized by the introduction of magnetic resonance imaging (MRI) with intravenous gadolinium-DTPA. The imaging technique has largely replaced myelography and radionuclide studies as a diagnostic tool. Radionuclear scans may still be indicated when an infectious or inflammatory process is in question. Plain radiographs and high-resolution computerized tomographs (CT) are still essential to the understanding of the osseous anatomy of the vertebral canal.

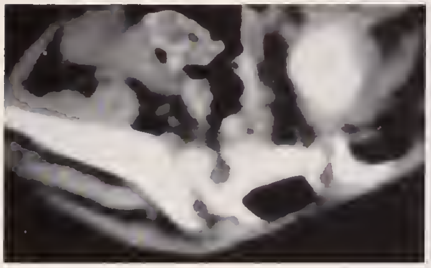
## Case Reports

Case 1. A 14-year-old girl had complaints of localized low back pain that occurred during athletic activity. Plain x-rays (Fig. 1) demonstrated a congenital block vertebra of the lumbar spine. MRI (Fig. 1) failed to reveal any neural compression. Pain was attributed to abnormal stresses placed on an immobile spine. With conservative management, she was able to participate in limited sporting events.

Case 2. A 4-year-old boy had progressive low back, left buttock, and leg pain. Plain x-rays (Fig. 2) revealed abnormal curvature of the spine as well as lumbosacral dysraphism. CT (Fig. 2) further showed the extent of the open neural arch and the presence of lipomatous tissue within the vertebral canal. MRI (Fig. 2) demonstrated abnormal lipomatous tissue extending to a tethered, low-lying spinal cord.

Case 3. An 11-year-old girl had low back pain and an elevated erythrocyte sedimentation rate. Plain x-rays (Fig. 3) demonstrated collapse





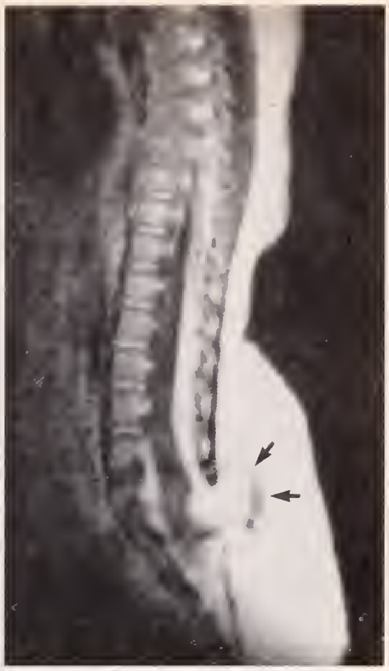
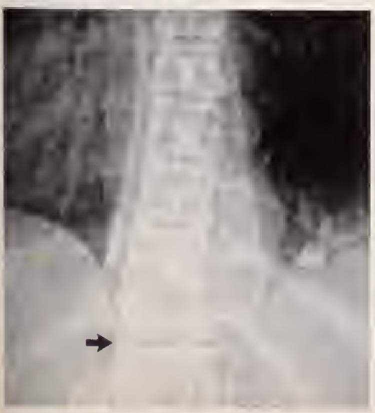
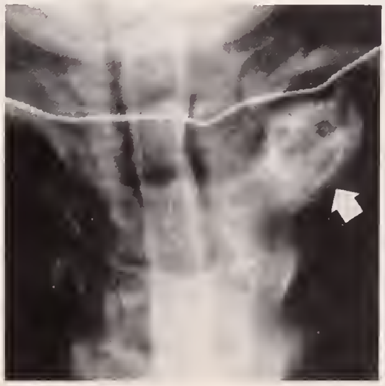
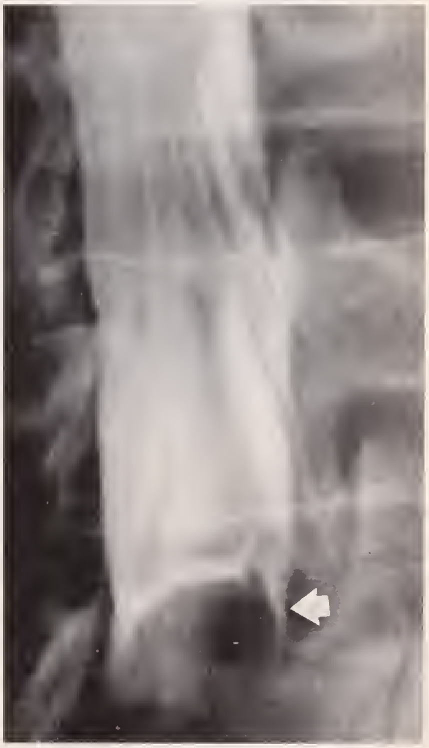


**Fig. 2.** Case 2. **Left** X-ray of lumbosacral spine showing scoliosis and dysraphism. **Middle** CT demonstrating defect in neural arch and lipoma in vertebral canal. **Right** MRI in sagittal plane revealing extent of lipomatous tissue (arrows) and tethered spinal cord.

**Fig. 3.** Case 3. **(Left)** X-ray of thoracic spine revealing compression of lower vertebral bodies (arrow).
**Fig. 4.** Case 4. **(Middle)** Iodopaque myelogram of thoracic spine demonstrating tumor of T1 vertebra (arrow).
**Fig. 5.** Case 5. **(Right)** Iodopaque lumbar myelogram showing complete block at L4–5 (arrow).







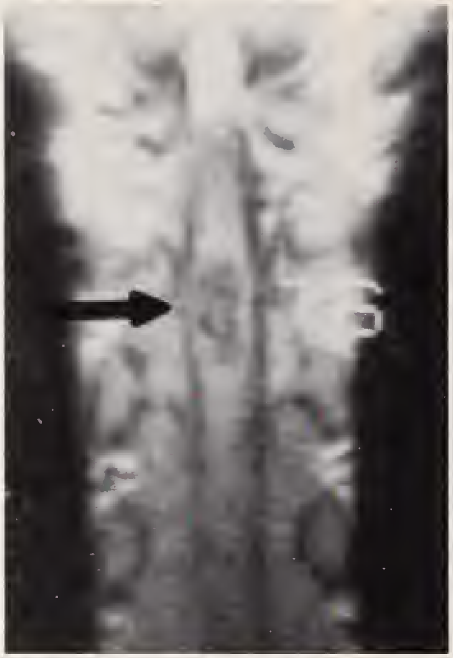
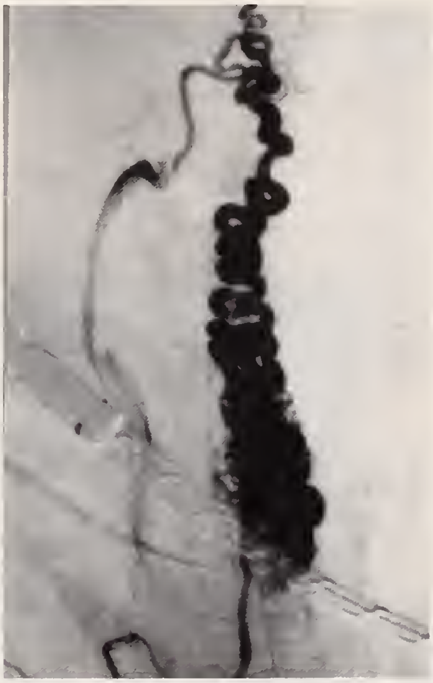


Fig. 6. (**Above**) MRI of thoracic cord demonstrating signal void at T6 (arrow). (**Below**) Selective angiograph showing glomus-type arteriovenous malformation supplied by T6 medullary artery.



of the lower thoracic vertebral bodies. Percutaneous aspiration revealed tubercle bacilli; successful treatment consisted of a combination of antituberculous agents and external bracing.

**Case 4.** A 17-year-old boy had back pain radiating down to the natal cleft. Myelographs (Fig. 4) demonstrated a mass arising from the pedicle and transverse process of T1. An osteochondroma was successfully resected by costotransversectomy.

**Case 5.** An 18-year-old boy with chronic low back pain developed right foot drop. There were no cutaneous stigmata of neurofibromatosis. Myelographs (Fig. 5) demonstrated a complete intradural block to flow of contrast medium. At the time of laminectomy, a solitary neurofibroma was discovered and resected with preservation of the L5 nerve root of origin. The pain and motor loss resolved completely.

## Discussion

**Congenital Causes.** Spondylolisthesis of either the dysplastic or isthmic type may be responsible for low back pain, emanating from the L5–S1 level in children (11). Chronic pain is frequently precipitated by an inciting traumatic event.

Initial treatment by external bracing followed by open fusion has been favored by many neurosurgeons. The presence of pseudoarthrosis at the defect in the pars interarticularis can be detected by MRI with gadolinium enhancement, and the distorted anatomy of exiting nerve roots due to reactive scarring can be demonstrated. Operative therapy at The Mount Sinai Hospital consists of removal of the free-floating posterior element, including the ascending facet, and resection of the pseudoarthrosis. No patient in our series has demonstrated further slippage.

Congenital block vertebra, a manifestion of Klippel-Feil syndrome, generally involves the cervical region. We have observed a block vertebra of the lumbar spine in a child with low back pain.

Lumbosacral leptomyelomeningocele is the most common lesion associated with the spinal dysraphic states (12). With advancing age, the history is often progressive motor, sensory, and sphincter deficits, as well as progressive orthopedic deformities. Pain and the presence of a subcutaneous mass over the lumbosacral spine and buttocks are the predominant features of the clinical syndrome.

We have previously outlined treatment emphasizing the disconnection of the subcutaneous lipomatous mass from the intradural neural elements and resection of the filum terminale to release the congenital tethered spinal cord (12–14). MRI is the imaging study of choice to visualize the extent of the lipoma and the traumatized, low-lying conus medullaris.

**Inflammatory and Infectious Causes.** Spontaneous diskitis and spinal osteomyelitis in children may be secondary to hematogenous spread from a staphylococcal focus (15). Diskitis in children most commonly occurs between the ages of one and twelve (mean age, 6 years). Erythrocyte sedimentation rate and a bone scan are the best methods of screening for an infectious process in a child with spinal pain and back tenderness. However, MRI enhanced with gadolinium-DTPA is rapidly coming into use.

Skeletal tuberculosis involving the vertebral column appears to predominate in children. Although previously associated with poor socioeconomic conditions, a resurgence has been intimately associated with the increase in the HIV-afflicted population (16, 17). The mainstay of treatment is medical therapy; radical debridement and open fusion may become necessary when antibiotic therapy fails.

**Neoplastic Origins.** Intramedullary tumors (astrocytoma, ependymoma, hemangioblastoma) and intradural tumors (Schwannomas) compose the majority of tumors in the child's spine (18). The osseous vertebral column is responsible for the remainder of these lesions (16).

Half of the reported cases of tumors of the spinal canal occur within the first year of life (5).

Either peripheral (type 1) or central (type 2) neurofibromatosis may initially appear as spinal manifestations in children, including abnormalities of the osseous structures and neurofibromas of the posterior roots.

Successful treatment of spinal tumors in children may require a multidisciplinary approach, including the use of laminoplasties and other stabilization techniques over areas of extensive bone removal in order to prevent late instability.

**Vascular Causes.** In earlier reports (19–21), the demographic distribution of the subtypes of vascular malformations of the spinal cord were clarified. Dural malformations of the spine in middle-aged persons do not have a tendency to hemorrhage. Children generally manifest the glomus and juvenile types of arteriovenous malformations, which have a great propensity to bleed from dilated arterial and venous participants. Apocalyptic back pain due to hemorrhage may be superimposed on a history of chronic back pain. Although MRI is the screening test of choice, angiography is still required to determine the therapeutic approach—embolization, surgery, or both (Fig. 6).

**Traumatic Causes.** The greater elasticity of the child's spinal column predisposes children to a high incidence of spinal cord injury without radiographic abnormality. However, those children whose radiographs evidence ligamentous disruption, osseous fracture, or misalignment have a high propensity for painful syndromes. In children, the traumatic incident may be trivial enough to be forgotten, and the spinal abnormality may only be discovered later (22).

## Summary

- Complaints of low back pain are rare in children
- Low back pain in children is not commonly related to degenerative or diskogenic disease
- Low back pain in children is not a manifestation of a psychiatric disorder
- Modern neuroradiologic evaluation is mandatory in any child with low back pain or spinal deformity
- Low back pain in children is frequently a harbinger of serious underlying disorders

## References

1. North RB. Acute back pain and disc herniation. Current therapy neurolog dis 1990; 3:57–60.
2. Mierau DR, Cassidy JD, Hamin T, Milne RA. Sacroiliac joint dysfunction and low back pain in school aged children. J Manipulative Physiolog Therapeutics 1984; 7:81–84.
3. Nepomuceno C, Faught E, Langford K. Diagnosis and management of low back pain. Alabama J Med Sci 1988; 25:430–435.
4. Hoffman HJ. Childhood and adolescent lumbar pain: differential diagnosis and management. Clin Neurosurg 1980; 27:553–576.
5. Miller JAA, Schmatiz C, Schultz AB. Lumbar disc degeneration: correlation with age, sex, and spine level in 600 autopsy specimens. Spine 1988; 13:173–178.
6. Nelson CL, Janecki CJ, Gildenberg PL, Sava G. Disc protrusions in the young. Clin Orthoped Rel Res 1972; 88:142–150.
7. Paajanen H, Erkintalo M, Kuusela T, Dahlstroms S, Kormano M. Magnetic resonance study of disc degeneration in young low-back patients. Spine 1989; 14:982–985.
8. Tertti M, Paajanen H, Kujala UM, Alanen A, Sami TT, Kormano M. Disc degeneration in young gymnasts: A magnetic resonance imaging study. Am J Sports Med 1990; 18:206–208.
9. Bunnell WP. Back pain in children. Orthoped Clin N Am 1982; 13:587–604.
10. Turner PG, Green JH, Galasko CSB. Back pain in childhood. Spine 1989; 14:812–814.
11. Wiltse LL, Jackson DW. Treatment of spondylolisthesis and spondylolysis in children. Clin Orthoped Rel Res 1976; 117:92–100.
12. Harrison MJ, Mitnick R, Rosenblum BR, Rothman AS. Leptomyelolipoma: analysis of twenty cases. J Neurosurg 1990; 73:360–367.
13. Rosenblum BR, Harrison MJ, Rothman AS. Lumbosacral lipomas. Contemp Neurosurg 1988; 10:1–8.
14. Hoffman HJ, Hendrick EB, Humphreys RP. The tethered spinal cord: its protean manifestations, diagnosis and surgical correction. Child's Brain 1976; 2:145–150.

15. Wenger DR, Bobechko WP, Gilda DL. The spectrum of intervertebral disc-space infection in children. JBJS 1979; 60A:100–108.

16. Camins MB, Rosenblum BR. Osseous lesions of the vertebral axis. In: Lewis MM, ed. Musculoskeletal oncology. Philadelphia, PA: WB Saunders (in press).

17. Ellinas PA, Rosner E. Pott's disease in urban populations: a report of five cases and a review of the literature. NY State J Med 1990; 90:588–591.

18. Tachdjian MO, Matson DD. Orthopedic aspects of intraspinal tumors in infants and children. JBJS 1965; 17A:223–248.

19. Rosenblum B, Oldfield E, Doppman J. Pathogenesis of spinal arteriovenous malformations. Surg Forum 1986; 37:489–491.

20. Rosenblum B, Oldfield EH, Doppman JI, DiChiro G. Spinal arteriovenous malformations: a comparison of dural arteriovenous fistulas and intradural AVMs in 81 patients. J Neurosurg 1987; 67:795–802.

21. Rosenblum B. Current concepts of spinal arteriovenous malformations. In: Rosenblum BR, ed. State of the art reviews in neurosurgery: Cerebral and spinal arteriovenous malformations. Philadelphia, PA: Hanley and Belfus, 1988, 1–12.

22. Rosenblum B, Rothman A. Cervical spine injuries in children. Ped Emergency News (in press).

# Lumbar Stenosis

Leonard I. Malis, M.D.

## Abstract

Lumbar stenosis continues to be inadequately appreciated as a major cause of sciatica, neurogenic claudication, and cauda equina syndrome. Review of the nineteenth-century medical literature reveals articles on narrow canal syndrome in achondroplastic dwarfs; connection to the clinical presentation of developmental spinal stenosis awaited the middle decades of the twentieth century. Appropriate neurosurgical intervention has evolved with a better understanding of the pathophysiologic changes that take place with this disorder and the introduction of myelography, computerized tomography, and magnetic resonance imaging.

Narrow lumbar canal syndrome remains an entity incompletely understood by most internists, neurologists, and orthopedists and many neurosurgeons. Although the medical literature from the nineteenth century (1–3) contains articles on lumbar stenosis, and appropriate surgical intervention (4) was available as early as 1900, full comprehension of the sequelae did not come about until the middle of the twentieth century. Imaging for diagnosis of a narrow lumbar canal became widely accessible only with the introduction of full-column myelography (5).

Between 1949 and 1990, Verbiest (6–15) published a series of articles on idiopathic lumbar spine stenosis associated with intermittent neurogenic claudication. In 1803, Portal (1) described vicious deviations of the spine, swelling of the vertebral bodies, and irregular outgrowth of the inner side of the laminae, producing considerable narrowing of the lumbar vertebral canal. Although Jaccoud (2) discussed cases of overly narrow canals compressing the spinal cord in 1864, not until 1910 did Sumita (16) identify congenital spinal stenosis as a cause of cauda equina syndrome in patients with chondrodystrophia fetalis, later termed achondroplasia by Parrot (3).

Decompressive laminectomy for relief of nerve root compression due to diffuse osseous overgrowth was performed by Sachs and Fraenkel (4) in 1900 and Bailey and Casamajor (17) in 1911; the findings at surgery included hypertrophic laminae, enlarged facet joints, and thickened ligamenta flava. Elsberg (18) in 1913 and Kennedy et al. (19) in 1914 reported a peculiar undescribed disease of the nerves of the cauda equina that today would be recognized as narrow canal syndrome. Farfan (20) later emphasized osteoarthritic changes in the three-joint complex consisting of the two posterior facets as well as the anterior disk at each level of the spine; spondyloarthrosis and degenerative disk disease coexist in acquired spinal stenosis.

The full-column technique of myelography (5) demonstrated by the late 1940s that in a stenotic patient in the standing position, the lumbar canal would fill up over the L1 level with only 6 cc of contrast media, whereas more than 12 cc was required in the full-sized normal canal (Fig. 1). There were washboard ridges at every level, and lateral constrictions as well. In attempting lumbar tap for a myelogram, difficulty in proper placement of the spinal needle within the thecal sac was highly suggestive; and cisternal puncture to inject Pantopaque was just beginning. Leakage into the subdural and epidural spaces was common. The spinous processes were often thickened and touched each other. Kissing facets occurred when the posterior articulations were

---

From the Department of Neurosurgery, Mount Sinai School of Medicine and The Mount Sinai Hospital, New York, NY. Address reprint requests to the author at 1048 Fifth Avenue, New York, NY 10029.

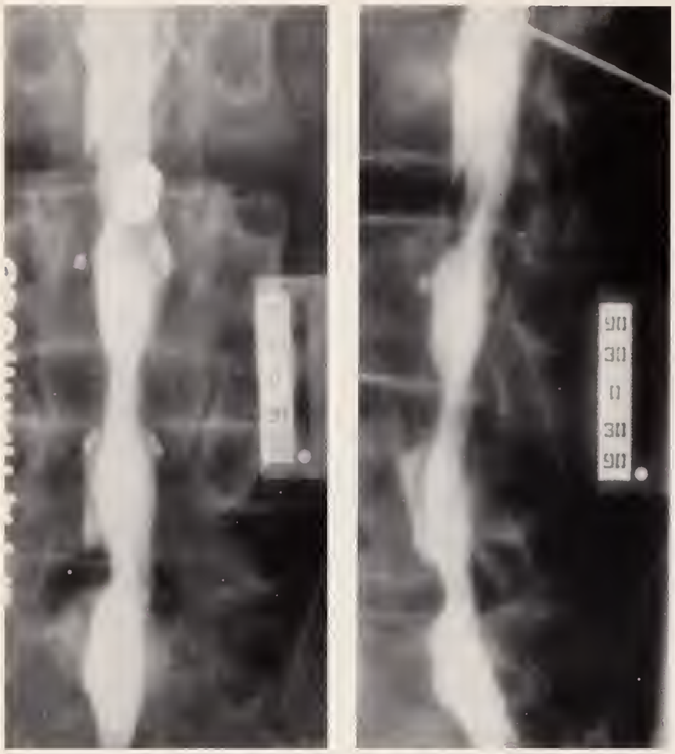


Fig. 1. Lumbar myelogram with Pantopaque; patient positioned at 90° (see scale). **Left** Note filling of entire canal by 6 cc of contrast media. **Right** Filling defects due to multiple disk protrusions.

huge and in close proximity to the spinous processes. Ventral spondylotic bars, hypertrophy of the ligamentum flavum, and compromised neural foramina further defined the neural compression.

In order to get an appreciation of canal stenosis, plain films of the spine were of far less value in the lumbar area than in the cervical spine. In the cervical region, the degree of stenosis correlated very well with the anterior-posterior diameter of the canal. However, in the lumbar spine, both the anterior-posterior diameter and interpediculate distance could be normal, yet—because of a trefoil configuration—the actual cross-sectional area could be tremendously constricted. Imaging of this abnormality was possible with multiple-slice lateral polytomography. The anterior-posterior diameter was measured at each level on every slice and then plotted on graph paper on the vertical axis; the horizontal axis corresponded to the distance between the tomographic cuts. This method of outlining gave a quite decent representation of what would have been an axial projection (Fig. 2). The normal range of the sagittal diameter of the lumbar spinal canal is 15 to 25 mm; a diameter less than 12 mm confirms stenosis. The range for normal interpediculate distance at the levels of L3 and L4 is 26 to 30 mm; a distance less than 24 mm suggests a narrow canal. Bulky, unwieldly, and expensive axial x-ray tomograph machines were later introduced; the films, while startling, were not nearly as sharp or detailed as our graph paper and pencil reconstructions.

Epstein and I reported in 1955 on the treatment of a number of achondroplastic dwarfs who required decompression of the entire spinal canal (21). With the patient in the knee-chest position, laminectomy was performed from the sacrum to the midthoracic area. When necessary, a second-stage operation was carried out from the midthoracic area to the foramen magnum with the patient in the sitting position. Only en face technique with the use of a Leksell rongeur was allowed to accomplish a midline laminectomy without any pressure on the dura before Kerrison punches were employed to widen the canal and

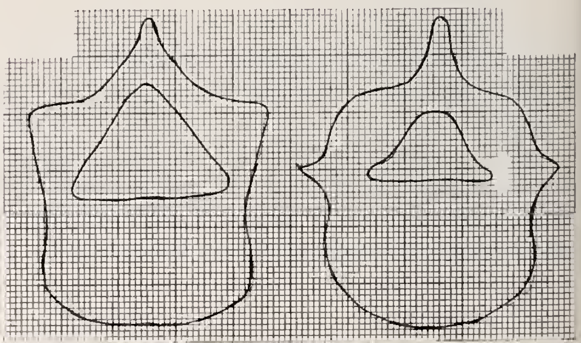
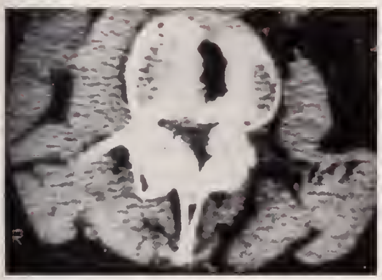


Fig. 2. Hand drawings of axial spine on 1-mm graph paper. **Left** Note normal canal with adequate dimensions. **Right** Spinal stenosis with shortened anteroposterior diameter and narrow interpediculate distance.



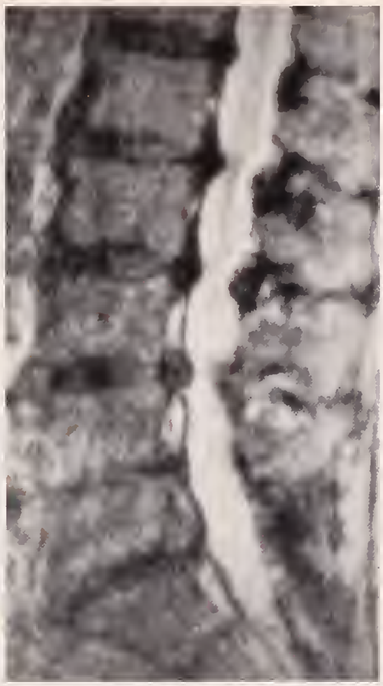


**Fig. 3. Left** High-resolution computed tomography of lumbar spine allows accurate measurement of axial spinal canal dimensions. **Fig. 4. Right** Magnetic resonance imaging, 2.0 Tesla, shows sagittal appearance of narrow canal with hypertrophic ligamenta flava and degenerative disk disease.

decompress the neural elements. Strict attention was paid to the foramina and lateral recesses. Regular admissions to the neurosurgical service at The Mount Sinai Hospital during the decades 1950 to 1970 included patients who had undergone diskectomy at several levels for herniated nucleus pulposus without attention to their narrow lumbar canals and without relief of their symptoms; following decompressive laminectomy, the cauda equina syndrome resolved in most of these patients.

Additional articles by Epstein (22, 23) added to the understanding of lumbar spine stenosis and lateral recess syndrome. The shape of the spinal canal varies; even when the ventrodorsal diameter is within normal limits, there may be a trefoil pattern with shortened height of the lateral recesses. Hypertrophy of bone in idiopathic developmental stenosis involves the structures of the vertebral arch: the pedicles, zygapophyseal joints, and laminae; the interlaminar spaces may be obliterated. The superior articular facets often extend or roof over the suprajacent intervertebral disc. Posterior osteophytes and subligamentous herniation of small amounts of nucleus pulposus may contribute considerably to nerve root compression within a small canal.

Neurogenic claudication associated with narrow canal syndrome mimics vascular claudication resulting from aortoiliac occlusive disease; peripheral neuropathy may also be present. Clinical features include intense dysesthesias and numbness in the lower limbs when walking or standing in any position that accentuates lumbar lordosis. The pain does not follow a radicular pattern. Symptoms disappear on sitting or lying down. When the patient changes posture or bends forward, the canal assumes a more generous dimension and relieves the pressure on neural elements.

The advent of computed tomographic scanning (Fig. 3) allowed imaging of the exact shape and dimensions of the bony spinal canal. Magnetic resonance imaging (Fig. 4) does not allow precise measurements of the canal; evidence of stenosis may be seen: multiple areas of bulging disks, hypertrophic ligamenta flava, large osteophytic ridges, and narrowness of the dural sac. The two neuroradiologic studies, when performed on high-resolution and 2.0 Tesla machines, are highly complementary and allow thorough preoperative evaluation.

Now that the clinical syndrome is well understood and diagnostic imaging is widely available, the operating neurosurgeon must still hone his or her surgical judgment. Developmental stenosis alone seldom produces low back pain. Unremitting sciatica may result from the addition of focal disk protrusion without frank herniation or from modest osteoarthritic changes with facet hypertrophy and vertebral body ridging. In such a case, laminotomy or laminectomy and foraminotomy become essential to avoid failed-back-surgery syndrome (24). On the other hand, identification of a narrow lumbar canal does not automatically dictate extensive surgical decompression. The patient with congenital or acquired stenosis who has a clear history of acute rupture of a nucleus pulposus may obtain permanent relief by microsurgical diskectomy.

Once the intervertebral disk had clearly been identified as the source of a herniated nucleus pulposus in 1934 (25), lumbar spinal stenosis remained inadequately appreciated as a separate cause of sciatica and neurogenic claudication. The earlier nineteenth-century recognition of narrow canal syndrome in achondroplastic dwarfism was not connected to developmental stenosis until after World War II. Appropriate neurosurgical intervention depends on full understanding of the pathophysiologic changes that take place in spinal stenosis and on proper interpretation of the preoperative neuroradiologic studies.

### References

1. Portal A. Cours d'anatomie médicale ou Elemens de l'anatomie de l'homme, 5 vols. Paris: Baudoin, 1803; vol. 1:299.
2. Jaccoud S. Les paraplegies et l'ataxie du mouvement. Paris: Adrien Delahaye, 1864.
3. Parrot MJ. Sur la malformation achondroplastique et le Dieu Phtah. Bull Soc Anthropol 1878; 1:296.
4. Sachs B, Fraenkel J. Progressive ankylotic rigidity of the spine (spondylose rhizomelique). J Nerv Ment Dis 1900; 27:1–15.
5. Malis LI, Newman C, Wolf BS. Full column technique of lumbar myelography. Radiology 1953; 60:18–28.
6. Verbiest H. Sur certaines formes rares de compression de la queue de cheval: 1. Les stenoses osseuses du canal vertebral. Paris: Hommage, Clovis Vincent, 1949.
7. Verbiest H. A radicular syndrome from developmental narrowing of the lumbar vertebral canal. J Bone Joint Surg 1954; 36B:230–237.
8. Verbiest H. Further experiences on the pathological influence of a developmental narrowness of the bony lumbar vertebral canal. J Bone Joint Surg 1955; 37B:576–583.
9. Verbiest H. Neurogenic intermittent claudication in cases of absolute and relative stenosis of the lumbar vertebra, in cases with narrow lumbar intervertebral foramina, and in cases with both entities. Clin Neurosurg 1973; 20:204–214.
10. Verbiest H. Pathomorphologic aspects of developmental lumbar stenosis. Orthop Clin North Am 1975; 6:177–196.
11. Verbiest H. Results of surgical treatment of idiopathic developmental stenosis of the lumbar vertebral canal: a review of twenty-seven years experience. J Bone Joint Surg 1977; 59B:181–188.
12. Verbiest H. Fallacies of the present definition, nomenclature and classification of the stenoses of the lumbar vertebral canal. Spine 1976; 1:217–225.
13. Verbiest H. The significance and principles of computerized axial tomography in idiopathic developmental stenosis of the bony lumbar vertebral canal. Spine 1979; 4:369–378.
14. Verbiest H. The treatment of lumbar spondyloptosis or impending lumbar spondyloptosis accompanied by neurologic deficit and/or neurogenic intermittent claudication. Spine 1979; 4:68–77.
15. Verbiest H. Stenosis of the lumbar vertebral canal and sciatica. Neurosurg Rev 1980; 3:75–89.
16. Sumita M. Beitrage zur Lehre von der Chondrodystrophia foetalis (Kaufmann) and Osteogenesis imperfecta (Vrolik) mit besonderer Berucksichtigung der anatomischen and klinischen Differential Diagnose. Dtsch Z Chir 1910; 107:1–110.
17. Bailey P, Casamajor L. Osteoarthritis of the spine as a cause of compression of the spinal cord and its roots. J Nerv Ment Dis 1911; 38:588–609.
18. Elsberg CA. Experiences in spinal surgery: observations upon 60 laminectomies for spinal disease. Surg Gynecol Obstet 1913; 16:117–132.
19. Kennedy F, Elsberg CA, Lambert CI. A peculiar undescribed disease of the nerves of the cauda equina. Am J Med Sci 1914; 147:645–667.
20. Farfan HF. Mechanical disorders of the low back. Philadelphia: Lee & Febiger, 1973.
21. Epstein JA, Malis LI. Compression of spinal cord and cauda equina in achondroplastic dwarfs. Neurology 1955; 5:875–881.
22. Epstein JA, Epstein BS, Lavine L. Nerve root compression associated with narrowing of the lumbar spinal canal. J Neurol Neurosurg Psychiatry 1962; 25:165–176.
23. Epstein JA, Epstein BS, Lavine LS, Carras R, Rosenthal AD, Sumner P. Lumbar nerve root compression at the intervertebral foramina caused by arthritis of the posterior facets. J Neurosurg 1973; 39:362–369.
24. Burton CV. How to avoid the "failed back surgery syndrome." In: Cauthen JC, ed. Lumbar spine surgery. Baltimore: Williams & Wilkins, 1982; 204–215.
25. Mixter WJ, Barr JS. Rupture of the intervertebral disc with involvement of the spinal canal. N Engl J Med 1934; 211:210–215.

# Synovial Cysts of the Lumbar Facet Joint

Ved P. Sachdev, M.D., Martin H. Savitz, M.D., Antoine I. Hindi, M.D., and Howard B. Goldstein, M.D.

## Abstract

A study of 42 synovial cysts arising from the lumbar facet joint is presented. Intraspinal cysts caused radicular pain and cauda equina syndrome in 11 patients. In a series of 1,400 lumbar laminotomies for a herniated disk, 31 incidental nodules arising from the exposed facet as paraarticular masses were found and excised. The limited number of reports in the orthopedic and neurosurgical literature has not dealt adequately with the etiology or incidence of these benign lesions. The neuroradiologic appearance and histopathologic findings are briefly discussed.

INTRASPINAL synovial cysts causing radicular pain, cauda equina syndrome, and myelopathy have been reviewed (1–4), most recently by Patel and Sanders (5). The reports have not dealt adequately with the etiology or incidence of these benign lesions. The histopathologic findings in synovial membranes following trauma are well known (6). Osteoarthritic changes in the facet joints are common in the presence of degenerative disk disease (7). Little mention has been made in the orthopedic or neurosurgical literature of hypertrophic synovitis of the zygapophyseal joint initially appearing as an incidental paraarticular mass (8–11).

## Case Reports

**Paraarticular mass.** In a series of 1,400 lumbar laminotomies, 31 synovial nodules were identified. During the opening, as the muscle was separated from the underlying spinous processes and laminae, a paraarticular mass with a narrow stalk arising from the exposed facet joint was uncovered (Fig. 1). Each of the encapsulated masses was approximately 1.0–1.5 cm in length and 3–5 mm in diameter. The first nodule was biopsied for frozen section and then removed in pieces. As the facet was partially resected, the origin of the stalk was noted to be the zygapophyseal joint. Subsequently, in the other 30 cases, when the entity was better recognized, the entire benign pedunculated lesion was amputated from the joint capsule prior to entering the disk space. The level involved was L2–3 in 1 patient, L3–4 in 2 patients, L4–5 in 24 patients, and L5–S1 in 4 patients.

Each of the 31 patients initially had neurologic signs and symptoms of lumbar disk disease. Conventional X-rays, computed tomography (CT) scanning, and myelography or magnetic resonance imaging (MRI) were performed in every case. Segmental spondylosis and herniated nucleus pulposus were the only neuroradiologic findings. Hypertrophic synovitis of the facet joint forming a paraarticular mass was an incidental finding at surgery. Laminotomy confirmed the clinical diagnosis of a herniated intervertebral disk.

The microscopic sections revealed fibroconnective tissue with areas of old hemorrhage, articular cartilage, and synovial membranes. In the illustrated case, there were many villous clefts interweaving in a multinodular stroma with areas of hemosiderin deposits. The histologic features were diagnostic of hypertrophic synovitis (Fig. 2).

**Benign Intraspinal Tumor.** Eleven patients had signs and symptoms of radiculopathy or

---

From the Department of Neurosurgery, The Mount Sinai Hospital, New York, NY, and the Departments of Pathology, Good Samaritan and Nyack Hospitals, Rockland County, NY. Address reprint requests to Ved P. Sachdev, M.D., 1148 Fifth Ave., New York, NY 10028.

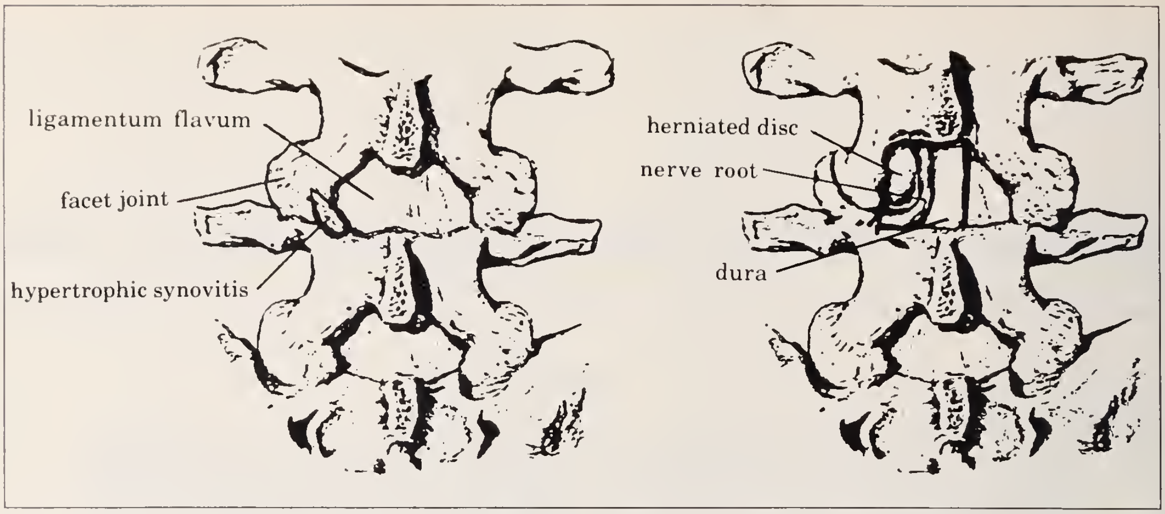


**Fig. 1.** **Left** Artist's drawing of operative findings. Note origin of pedunculated mass arising from zygapophyseal joint. **Right** Complete resection of lesion during laminotomy to expose herniated disk.

cauda equina syndrome. All of the mass lesions were located posterolaterally. The diagnosis was made preoperatively, and a decompressive laminectomy was performed in each case. The level involved was L3–4 in 1 patient, L4–5 in 9 patients, and L5–S1 in 1 patient.

The neuroradiologic studies were consistent with an epidural mass. In the first illustrated case (Fig. 3), the CT and MRI revealed a spherical intraspinal cyst in relation to the facet joint. At surgery, the nerve root was noted to be compressed by the 1.0-cm diameter lesion. In another case (Fig. 4), the CT and MRI suggested a large epidural mass almost completely obliterating the spinal canal. At surgery, the cyst was 2.0 × 3.0 cm, with adhesions to the dura and ligamentum flavum.

In the last illustrated case, microscopic sections revealed that the synovial cyst became confluent with the ligamentum flavum; higher-power views showed articular cartilage embedded in connective tissue and synovial clefts (Fig. 5). No neoplastic cells were observed.

## Discussion

Kao et al. (8) suggested that there were two types of juxtafacet cysts: a synovial cyst and a ganglion cyst, both arising from periarticular tissue. Theories of etiology revolved around at-

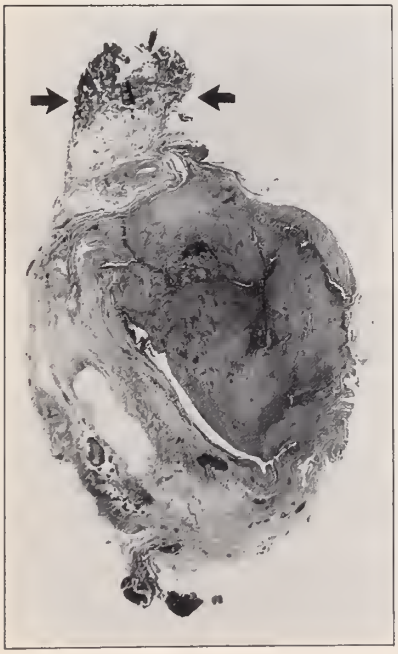


**Fig. 2.** Low-power, cross-sectional view of entire encapsulated nodule, showing multiple synovial sinuses with areas of hemosederin-laden phagocytes due to old hemorrhage. Note connective tissue stalk (arrows).

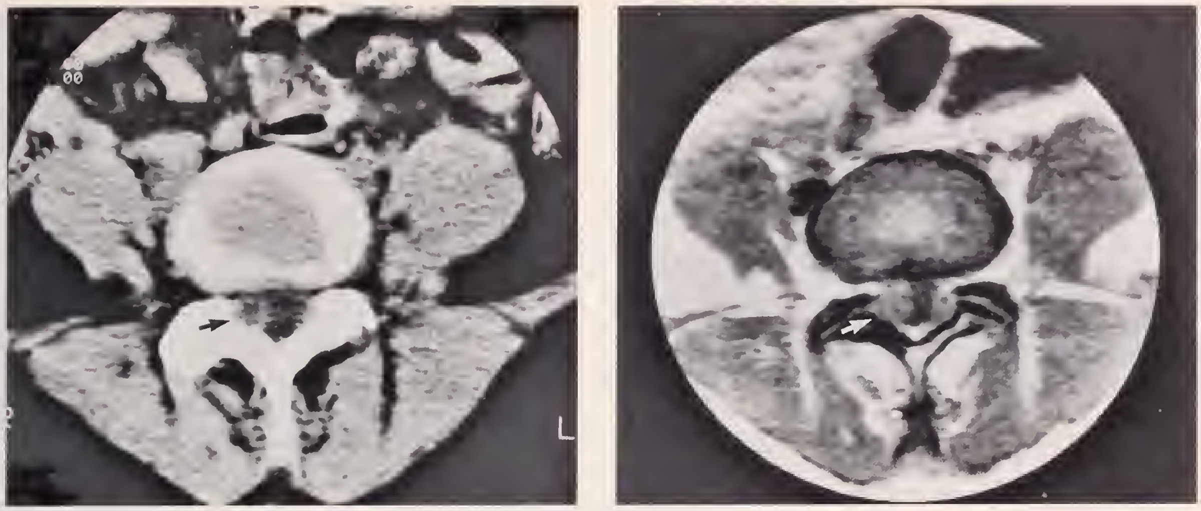


Fig. 3. **Left** CT of L4–5 level. Note cyst (arrow) arising from right apophyseal joint. **Right** Axial MRI confirming cystic lesion (arrow) originating in right L4–5 facet.

tempts to distinguish true synovial cysts from intraspinal ganglia and paraarticular nodules (1, 2, 12). Abdullah et al. (13) described 4 cases of synovial cysts arising from the ligamentum flavum, and a few ganglion cysts have been reported that contained gas adjacent to a vacuum zygapophyseal joint (14–17).

Degenerative disease in the intervertebral disk is generally associated with damage to the posterior joints. Minor strains of the articular ligaments produce synovitis of the facet joint. When the synovium is torn, hemarthrosis results (6). As in all diarthrodial joints, fibrillation and degeneration of cartilage occur with erosion; osteophytosis forms at the periphery, or the entire articular process may increase in size (18). Changes in the joint capsule accompany changes in the articular surfaces. The dorsal capsule thickens and frequently develops areas of cartilage within its substance. The articular cartilaginous ligaments may also become lax and redundant. With continuing avulsion, reinjury, and scarring, synovial loculations can form and trap inside fibrotic elements, connective tissue, or clefts representing collapsed synovium.

In studies of autopsy specimens, nodules considered primarily cartilaginous in origin have been described as varying in size from barely visible to large masses protruding like beads behind and below the facet articulations (19). No tumors

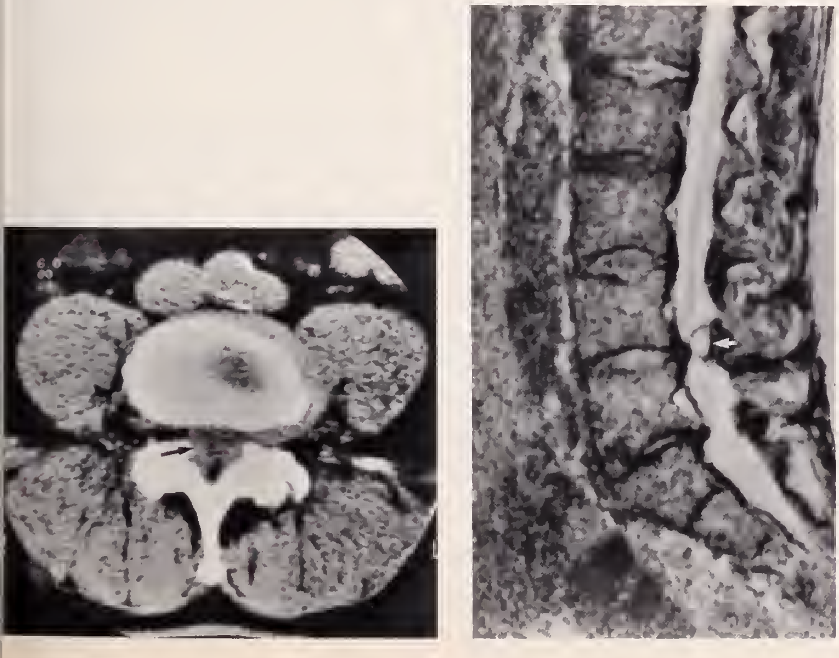
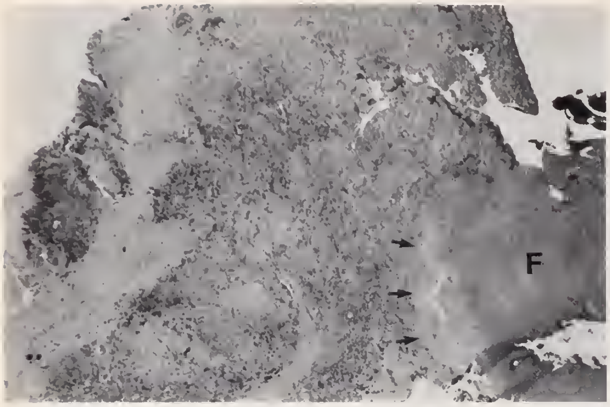


Fig. 4. **Left** CT of L4–5 intervertebral disk space. Note irregular mass lesion anterior to right facet joint (arrow). **Right** Sagittal T2 weighted MRI. Note epidural mass in posterior aspect of canal with bright signal intensity typical of a cyst (arrow) almost completely compromising lumbar theca.



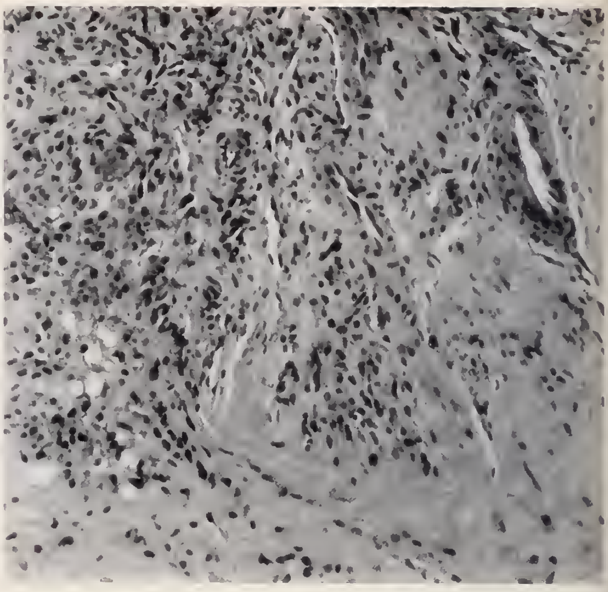


Fig. 5. **Left** Low-power view of one edge of large synovial cyst showing area of adherent (arrows) to adjacent ligamentum flavum (F). **Right** Higher-power section. Note multiple slit-like synovial clefts.

were identified in the neuroradiologic studies of the 31 cases with paraarticular masses because the lesions were not cystic, did not displace neural elements, and had a density similar to that of other extracapsular tissue. In the 11 patients with epidural masses visualized on CT and MRI, the synovial nodules arose medial to the joint and compressed one or more nerve roots. Several of the lesions attached to the ligamentum flavum and dura, but the true origin was the zygapophyseal joint. Hypertrophic synovitis arising from the facet is a common pathologic process which may be observed rarely as a true intraspinal neoplasm, but appears to occur incidentally as a paraarticular mass in 2% of lumbar laminotomy procedures for removal of a herniated disk.

## References

1. Baum JA, Hanley EN. Intraspinal synovial cyst simulating spinal stenosis. Spine 1986; 11:487–489.
2. Kjerulf TD, Terry DW, Boubelik RJ. Lumbar synovial or ganglion cyst. Neurosurgery 1986; 19:415–420.
3. Onofrio BM, Mih AD. Synovial cysts of the spine. Neurosurgery 1988; 22:642–647.
4. Pendleton B, Carl B, Pollay M. Spinal extradural benign synovial or ganglion cyst: case report and review of literature. Neurosurgery 1983; 13:322–325.
5. Patel SC, Sanders WP. Synovial cysts of the spine: case report and review of the literature. Am J Neuroradiol 1988; 9:602–603.
6. Soren A, Klein W, Huth I. The synovial changes in posttraumatic synovitis and osteoarthritis. Rheum Rehab 1978; 17:38–45.
7. Lewin T. Osteoarthritis in lumbar synovial joints. Acta Orthop Scand 1964; 73(Suppl).
8. Kao CC, Winkler SS, Turner JH. Synovial cyst of spinal facet: a case report. J Neurosurg 1974; 41:372–376.
9. Savitz MH, Katz SS, Goldstein H, Worcester D. Hypertrophic synovitis of the lumbar facet joint in two cases of herniated intervertebral disc. Mt Sinai J Med 1982; 49:434–437.
10. Savitz MH, Katz SS, Goldstein HB, Worcester D. Hypertrophic synovitis of the facet joint forming a paraarticular mass in cases of herniated intervertebral disc. Spine 1987; 2:509–510.
11. Savitz MH, Katz SS. Hypertrophic synovitis of the lumbar facet joint (letter). Neurosurgery 1988; 22:970–971.
12. Bhushan C, Hodges FJ, Wityk JJ. Synovial cyst(ganglion) of the lumbar spine simulating extradural mass. Neuroradiology 1979; 18:263–268.
13. Abdullah A, Chambers R, Daut D. Lumbar nerve root compression by synovial cysts of the ligamentum flavum. J Neurosurg 1985; 60:617–620.
14. Beatty RA. Sciatica and epidural gas. Neurosurgery 1987; 21:537–539.
15. Schulz E, West W, Hinshaw D, Johnson R. Gas in a lumbar extradural juxta-articular cyst: sign of synovial origin. Am J Radiol 1984; 143:875–876.
16. Silbergleit R, Gebarski SS, Brunberg JA, McGillicudy J, Blaivas M. Lumbar synovial cysts: correlation of myelographic, CT, MR, and pathologic findings. AJNR 1990; 11:777–779.
17. Spencer RR, Jahnke RW, Hardy TL. Dissection of gas into an intraspinal synovial cyst from contiguous vacuum facet. J Comput Assist Tomogr 1983; 7:886–888.
18. Helfet AJ, Lee DMG. Disorders of the lumbar spine. Philadelphia: JB Lippincott, 1978:27–30.
19. Harris RI, MacNab I. Structural changes in the lumbar intervertebral discs. J Bone Joint Surg 1954; 36B:304–322.

# Preoperative and Postoperative Neuroradiologic Evaluation of Lumbar Disk Herniation

Robert Tash, M.D., Richard C. Strauss, M.D., and Martin H. Savitz, M.D.

## Abstract

Three hundred and forty patients with signs and symptoms of lumbar disk disease were evaluated preoperatively by conventional X-rays, computed tomography, and 2.0 Tesla magnetic resonance imaging (MRI). High-resolution T2 weighted MRI provided a truly noninvasive myelogram. The operative findings were predicted by and correlated with the neuroradiologic studies. High-resolution MRI with intravenous gadolinium-DTPA (Magnevist, Berlex) has been particularly useful in evaluating patients with a history of prior disk surgery and patients who were not relieved of pain by lumbar diskectomy.

Three years' experience with 160 lumbar diskectomies following computed tomography (CT) and 0.3 Tesla magnetic resonance imaging (MRI) and without myelography was reported in 1988 (1). A subsequent series of 340 cases with signs and symptoms of lumbar disk disease has now been evaluated preoperatively by plain X-rays, CT, and 2.0 Tesla MRI. None of the patients underwent conventional myelography or CT myelography.

## Methods

All of the MRI studies were performed on a 2.0 Tesla superconductive magnet (Elscint). Sagittal images were obtained with a T1 weighted pulse sequence (TE 20 mS, TR 350 mS) as well as with proton density and T2 weighted pulse sequences (TE 20 mS and 81 mS, TR 2000 mS). Two acquisitions with a 192 × 256 matrix size were employed. The slice thickness for the majority of the images was 4 or 5 mm. In nonoperated-on patients, axial proton density and T2 weighted images (TE 29 mS and 80 mS, TR 2000 mS) were obtained. In postoperative cases, axial T1 images (TE 15.4 mS, TR 450 mS) were produced with and without the intravenous administration of gadolinium 0.1 mM/kg; sagittal images were formulated with the same parameters.

## Case Reports

Case 1. A 32-year-old man remained symptomatic with sciatica on the right side 6 months after a moving-vehicle accident. Neurologic examination revealed right-side paravertebral spasm with pelvic tilting, right-side L5 hypesthesia, and a depressed right ankle jerk. CT was consistent with a herniated nucleus pulposus at the right-side L4–5 level. MRI confirmed the disk herniation at L4–5 (Fig. 1), but also revealed an incidental herniated disk at the left L5–S1 level (Fig. 1). The patient underwent minilaminotomy only at the symptomatic level.

Case 2. A 45-year-old woman was immediately relieved of her symptoms after undergoing minilaminotomy at the right-side L4–5 level. Preoperative CT was performed and MRI (Fig. 2) confirmed the diagnosis. Three months after discharge, she experienced recurrence of right-sided sciatica. Repeat MRI (Fig. 2) showed some amor-

---

From MRI Diagnostics of Rockland, Inc., Suffern, NY, and the Department of Neurosurgery, The Mount Sinai Hospital, New York, NY. Address reprint requests to Robert Tash, M.D., MRI Diagnostics of Rockland, Inc., 11 North Airmont Road, Suffern, NY 10901.

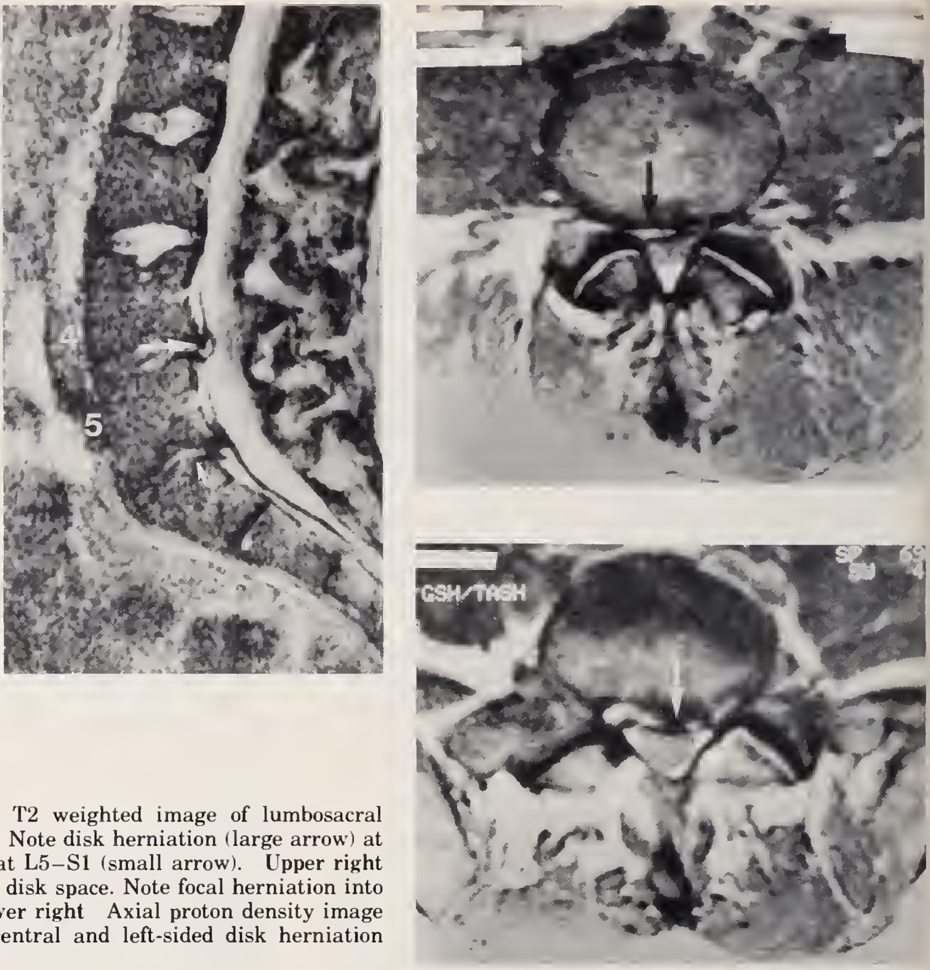


Fig. 1. Case 1. Left Sagittal T2 weighted image of lumbosacral spine, just to the right of midline. Note disk herniation (large arrow) at L4–5 and degenerative changes at L5–S1 (small arrow). Upper right Axial proton density image L4–5 disk space. Note focal herniation into right lateral recess (arrow). Lower right Axial proton density image L5–S1 disk space. Note small central and left-sided disk herniation (arrow).

phous material in the right intervertebral foramen; imaging with gadolinium (Fig. 2) revealed only enhancing scar tissue. The patient improved on a course of physiotherapy and antiinflammatory agent.

Case 3. A 62-year-old woman had undergone laminectomy and fusion at L4–5 and L5–S1. Six years later, she presented with left-sided sciatica and weakness of the quadriceps femoris. Preoperative CT and MRI demonstrated an L3–4 herniated nucleus pulposus. Following minilaminotomy at the left L3–4 level, her symptoms were not relieved. Repeat MRI with gadolinium (Fig. 3) one week after surgery revealed a postoperative seroma extending through the laminotomy site and well above the L3–4 level. A third

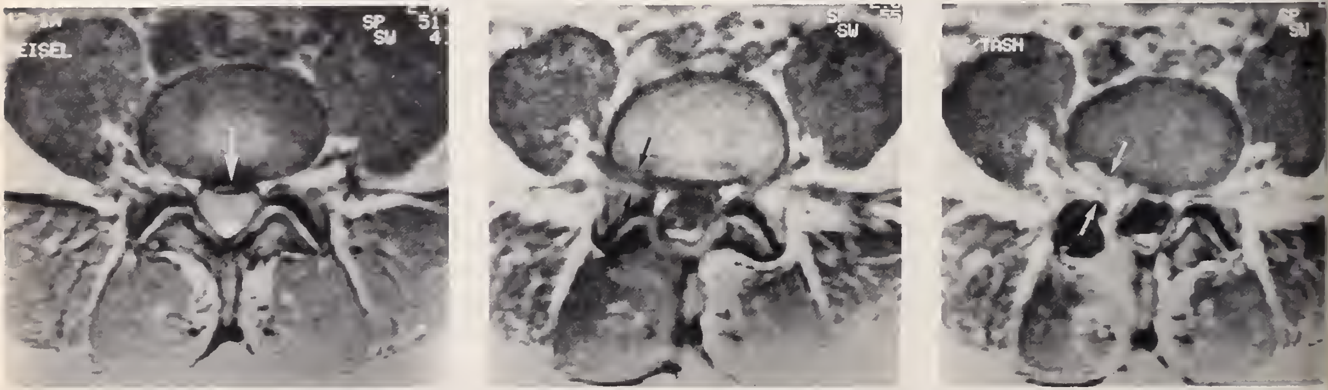
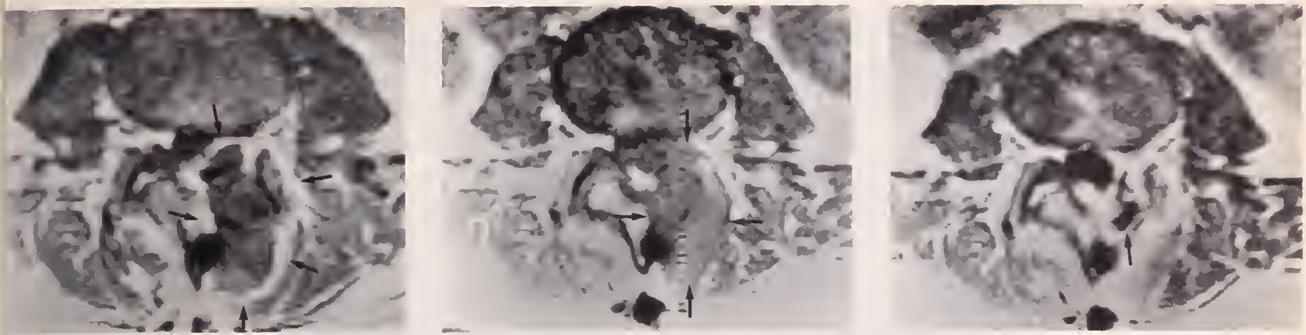


Fig. 2. Case 2. Left Axial proton density image L4–5 intervertebral disk. Note central herniation of disk material (arrow). M dle Axial T1 weighted image L4–5 disk space. Note right laminotomy and amorphous soft tissue in the right intervertebral foram (arrows). Right Following intravenous gadolinium, note enhancement of soft-tissue mass as well as posterior aspect of adjace anulus (arrows). Findings are consistent with epidural fibrosis.



ig. 3. Case 3. **Left** Axial T1 weighted image L3–4 intervertebral disk following intravenous gadolinium. Within left paraverte-ral musculature, note large mass with enhancing rim (arrows) extending through laminotomy site into posterior aspect of spinal anal. **Middle** Axial T1 weighted image L3–4 disk space. Note persistent soft-tissue mass at laminotomy site (arrows). **Right** ollowing intravenous gadolinium, note enhancement of majority of soft-tissue mass except for residual seroma in center (arrow).

MRI (Fig. 3) eight weeks later showed a persistent soft-tissue mass. Imaging with gadolinium (Fig. 3) revealed enhancing epidural scar causing significant distortion of the thecal sac. The patient underwent reexploration at L3–4; the findings at surgery were hypertrophic scar and ligamenta flava compressing the dura and the left L3 nerve root in the intervertebral foramen.

Case 4. A 55-year-old woman underwent microsurgical diskectomy at the L5–S1 level with good relief of her low back pain. Preoperative CT was consistent with an L5–S1 herniated nucleus pulposus and disk disease at L4–5 without frank herniation. Two months after surgery, she became symptomatic with left-sided sciatica. MRI at L5–S1 (Fig. 4) showed a soft-tissue mass just above the level of surgical exploration; imaging with gadolinium (Fig. 4) revealed enhancing epidural fibrosis with a central fragment of intervertebral disk. The findings at surgery for reexploration of L5–S1 were large amounts of granulation tissue in the lateral recess above the disk space and residual disk material.

Case 5. A 45-year-old woman had undergone laminectomy and diskectomy at the right-side L4–5 level without complication. Two years later, recurrent sciatica was evaluated by CT and MRI, which demonstrated scarring and segmental spondylosis at the L4–5 level. Following reexploration and lysis of adhesions, she developed severe unremitting lumbosacral spasms. The erythrocyte sedimentation rate remained markedly elevated. A repeat MRI study (Fig. 5) was consistent with diskitis. Three separate biopsies of the disk space failed to culture any organism. The patient was treated with eight weeks of bed rest and mobilized in a Knight spinal brace.

## Discussion

High-resolution MRI of the spine can characterize disk herniation with spatial resolution comparable to CT and myelography (2). As with CT, multiplanar MRI is suited for demonstrating both central and lateral disk abnormalities (3–5). As a truly noninvasive myelogram, 2.0

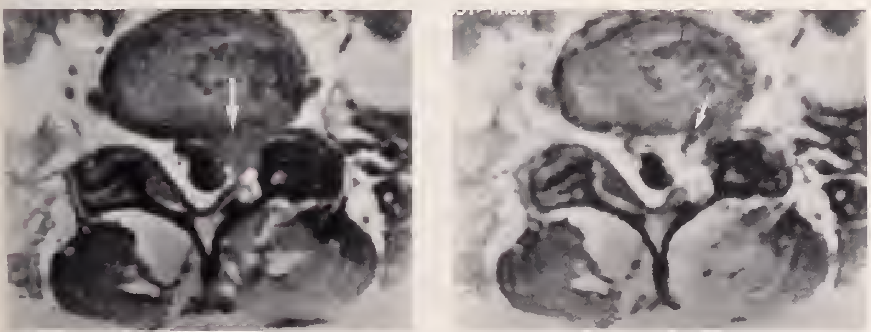


Fig. 4. Case 4. **Left** Axial T1 weighted image L5–S1 intervertebral disk. Note partial laminectomy and soft-tissue mass in left lateral recess (arrow). **Right** Following intravenous gadolinium, note enhancement of extradural scar with exception of central herniated disk material (arrow).

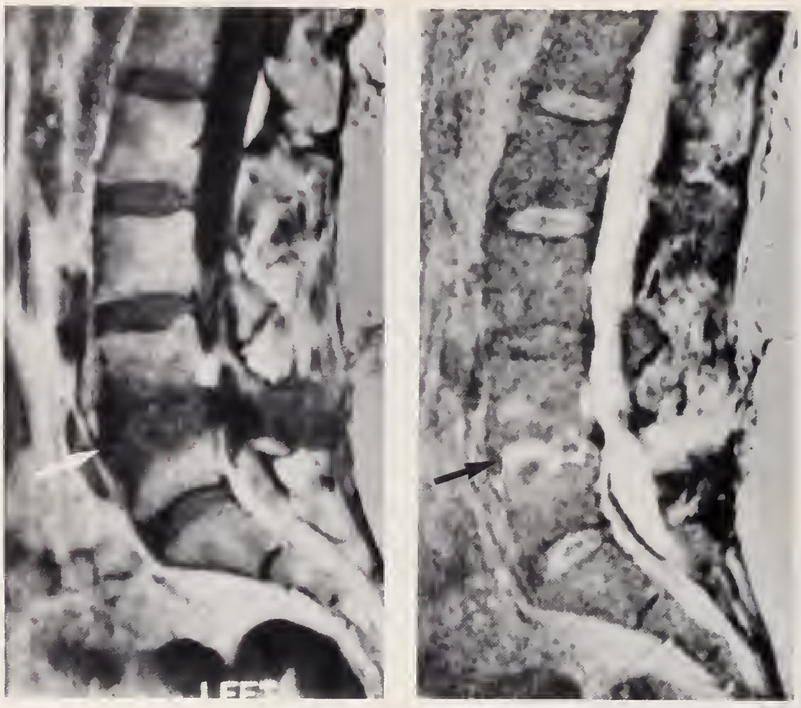


**Fig. 5.** Case 5. **Left** Sagittal T1 weighted image of lumbosacral spine. Note indistinct margins of end plates at L4–5 (arrow) and decrease in signal intensity of adjacent subchondral bone. **Right** Sagittal T2 weighted image. Note increase in signal intensity of L4–5 intervertebral disk (arrow). Again demonstrated are irregularities of both end plates.

Tesla MRI has the capacity to image the entire spinal canal and thecal contents, not only in the sagittal plane, but in the axial and coronal planes as well. Side effects associated with the use of gadolinium-DTPA are rare when compared to the complications of other intravenous contrast media.

We were not willing to abandon CT for preoperative evaluation. Thinner section imaging is less subject to volume-averaging artifact. CT is better able to show degenerative bony changes along the vertebral end plates and facets and calcifications within the anulus fibrosus and herniated disk fragments.

## Summary

CT and MRI are ideally complementary examinations to evaluate both intradural and extradural abnormalities. The combination of plain X-rays, CT, and MRI is sufficient and complete for the preoperative diagnosis of patients with suspected lumbar disk disease and makes conventional myelography and CT myelography unnecessary. Intrathecal injection of contrast media should be reserved for those patients in whom other imaging modalities are equivocal or who have some contraindication to MRI, such as a pacemaker or an aneurysm clip.

High-resolution T2 weighted MRI yields a completely noninvasive myelogram. Intravenous gadolinium-DTPA in previously operated-on patients is highly successful in distinguishing scar tissue from recurrent disk herniation. The operative findings in a combined series of 500 patients who underwent minilaminotomy were predicted by and correlated with neuroradiologic studies.

### References

1. Savitz MH, Katz SS, Zablow BC. Discectomy following CT and MRI and without myelography. Mt Sinai J Med 1988; 55:162–166.
2. Edelman RR, Shoukimas GM, Stark DD, et al. High-resolution surface-coil imaging of lumbar disc disease. AJR 1985; 144:1123–1129.
3. Han JS, Kaufman B, El Yousef SJ. NMR imaging of the spine. AJR 1983; 4:1151–1159.
4. Modic MT, Weinstein MA, Pavlicek W. Nuclear magnetic resonance imaging of the spine. Radiology 1984; 148:757–762.
5. Modic MT, Pavlicek W, Weinstein MA. Magnetic resonance imaging of intervertebral disc disease: clinical and pulse sequence considerations. Radiology 1984; 152:103–111.

# Role of Laminectomy in Lumbar Disk Herniation

Charles A. Fager, M.D.

## Abstract

A personal view of the role of laminectomy in lumbar disk herniation is presented. The phenomena of the true herniated nucleus pulposus and the asymptomatic bulging degenerative disk are discussed in relation to clinical presentation, neuroradiologic studies, and surgical findings. Microscopic technique and so-called less invasive procedures are also evaluated.

Because no cure exists for lumbar disk disease, surgical treatment must be reserved for the patient with unmistakable evidence of neural compression, with the goals of relieving radicular pain and restoring neurologic function. Furthermore, in view of the fact that most patients with a protruded disk and even some patients with completely extruded disk fragments make a complete recovery without treatment (1, 2), surgical intervention is required only when it becomes clear that the patient will not otherwise recover. Naturally, earlier operation may be necessary in the exceptional circumstance of a patient whose neurologic condition is in serious jeopardy or who has incapacitating radicular pain that cannot be controlled by other measures.

When surgical treatment is necessary for lumbar disk disease with neural compression, it should follow the same well-established principles as any intraspinal surgery, including adequate decompressive exposure of the nerve root or the neural elements under compression, longitudinal release of the nerve root or cauda equina before excision of the disk, and a lateral approach to the lesion. Adherence to these principles provides clear documentation of all the pathologic conditions that produce neural compression, delineation of the ruptured fragment of disk regardless of its size or location, and atraumatic dissection within the spinal lumen. The approach for a disk lesion compressing the cauda equina should be the same as for any epidural mass. The difference in surgical technique for a smaller disk lesion should only be quantitative, ensuring relief of neural compression and prevention of operative neurologic deficit.

**Adequacy of Exposure.** Providing decompressive exposure should eliminate the possibility of unrelieved compression by a retained fragment of disk or by associated facet arthropathy. Adequate exposure is also effective in preventing nerve root damage, surgery at the wrong level, dural tear, symptomatic epidural fibrosis, and adhesive arachnoiditis. These complications have been observed in patients who have had operation with little or no excision of lamina for removal of disk through a small opening, sometimes in the presence of spinal stenosis (Fig. 1). Rather than relieving compression of the nerve root or cauda equina, such an operation has intensified constriction because of contusion, early root edema in a confined canal, later epidural fibrosis, and cohesion of nerve roots, often termed arachnoiditis.

**Microdiskectomy.** The procedure known as lumbar microdiskectomy has resulted in much smaller openings en route to the pathologic process causing the disk herniation. Although the operating microscope cannot be criticized as a

---

From the Department of Neurosurgery, Lahey Clinic Medical Center, Burlington, MA. Address reprint requests to Charles A. Fager, M.D., Department of Neurosurgery, Lahey Clinic Medical Center, 41 Mall Road, Burlington, MA 01805.

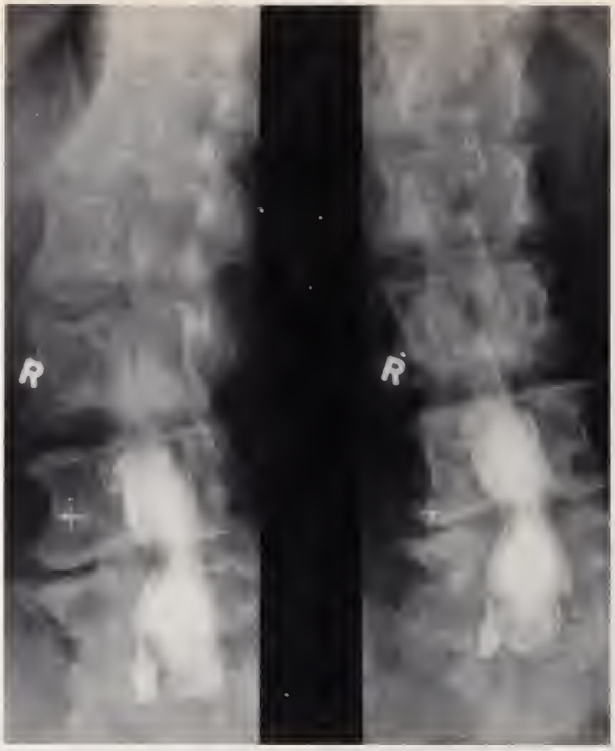


Fig. 1. Characteristic myelographic appearance of right L5 nerve root compression caused by facet arthropathy (lateral recess stenosis) often mistaken for protrusion of disk.

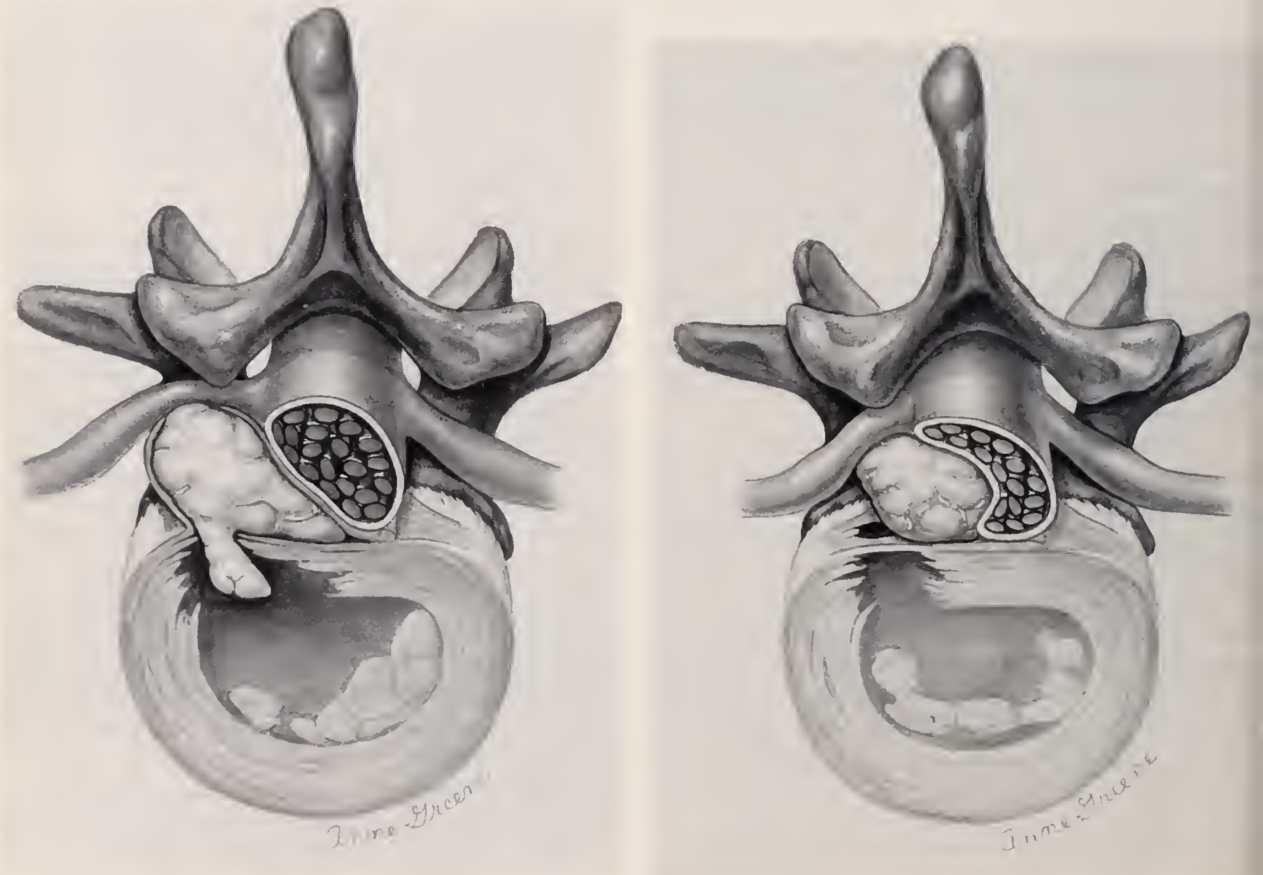


Fig. 2. Rupture and extrusion of disk through opening in anulus much smaller than fragment itself. Note small stem of disk within disk space but not contiguous with intradiskal tissue on left. *Reprinted* with permission from Fager CA. Lumbar micro-discectomy: a contrary opinion. Clin Neurosurg 1986; 33:422, 423.

useful tool in spinal surgery, it should not be employed in contradistinction to these principles, but rather to enhance the dissection when the surgeon believes that magnification is helpful. The microscope itself should not generate an additional surgical fee when used for lumbar disk surgery. Previous reports have enunciated the pitfalls and disadvantages of microdiskectomy (3) and enumerated the reasons for failure in patients who have required additional operations (4).

**Ruptured Disks.** The phenomenon of ruptured disk has intrigued and mystified those who confront such a finding. An observation of less than passing importance has been that the patient who requires surgery is often found at the operating table to have a fragment of disk, whether subligamentous or fully extruded, that has erupted into the spinal canal through an opening in the anulus fibrosus much smaller than the fragment itself (Fig. 2). It has literally been squeezed or pushed or has exploded through such a small hole in a dumbbell-like or mushroom display. The fragment is completely disrupted and is not in continuity with remaining disk material or other fragments within the disk space. The phenomenon is easy to understand because the major part of the offending fragment or head of the mushroom is torn free during propulsion. This type of biomechanical trajectory invalidates any intradiskal therapy.

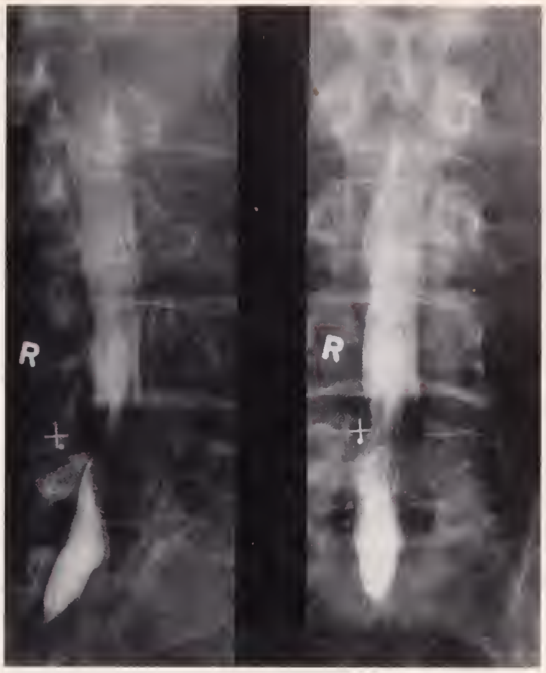


Fig. 3. Lumbar myelograms showing large disk fragment extruded cephalad from supramarginal tear in anulus at L5–S1. Note compression of left L5 and S1 nerve roots.

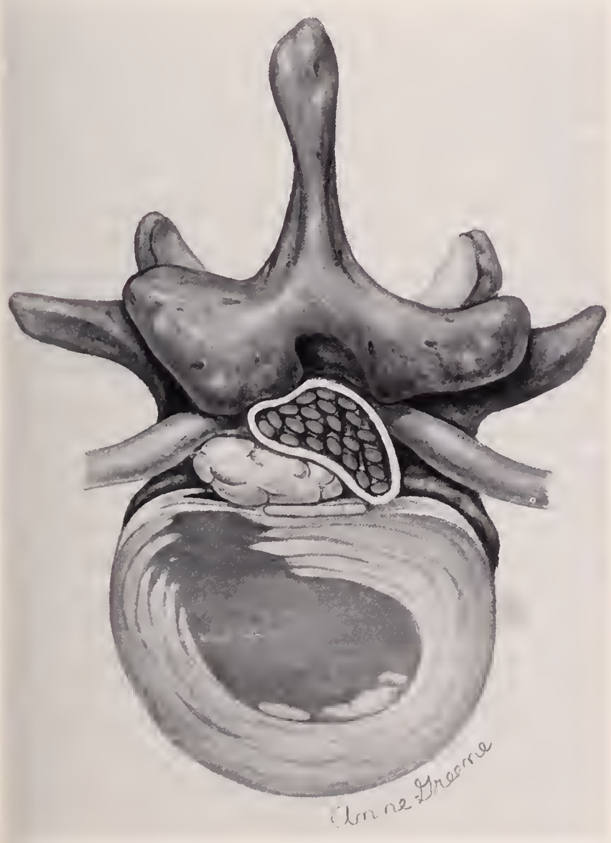


Fig. 4. Facet arthropathy compressing nerve root and cauda equina from posterolaterally and extruded disk from anteriorly. *Reprinted* with permission from Fager CA. Lumbar microdiscectomy: a contrary opinion. Clin Neurosurg 1986; 33:432.

Of similar magnitude is the fairly common observation that some disks rupture through a marginal tear in the anulus, an opening also smaller than the detached fragment itself. These disk fragments are displaced caudad or cephalad, a greater distance when fully extruded than when subligamentous (Fig. 3). They may impact a nerve root within its foramen above or below the disk space or compress the cauda equina when lying in a median or paramedian position.

Although intradiskal nucleolysis or aspiration will surely reduce elements of disk tissue confined within the disk space, the procedure cannot possibly lyse or retrieve fragments that have escaped into the spinal canal through a small aperture. Simple protrusion, or what has been referred to as “bulging,” of the disk may be amenable to such an approach, but most patients with this type of disk problem in which the anulus remains intact make a satisfactory recovery without any specific treatment.

Unfortunately, even normal prominence of a disk, especially at the L4–5 level, has been a tar-

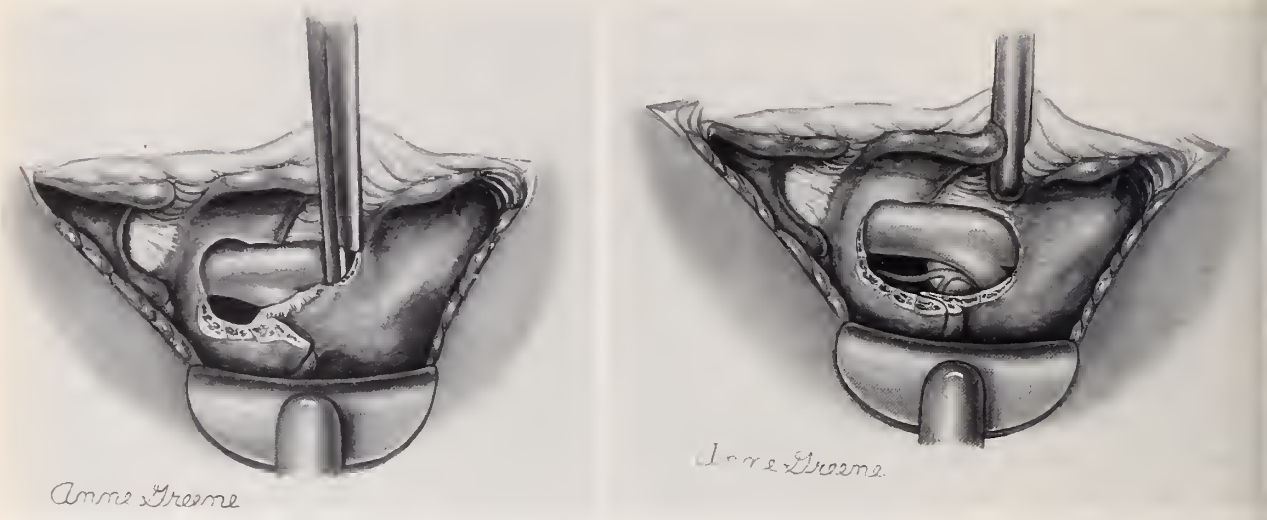


**Fig. 5.** **Left** Adequate removal of bone around compressed nerve root. **Right** Release of nerve root and lateral approach to fragment of disk. *Reprinted* with the permission of the Lahey Clinic.

get for those determined to pursue a "less invasive" treatment. This practice continues to be a problem that sometimes leads to inappropriate treatment when the indications for intervention are far from clear. Adding to the diagnostic dilemma, radiologists too frequently report protrusion of a disk when the prominence seen by myelography or imaging is simply associated with ordinary degeneration and narrowing of a disk space.

When surgery becomes necessary, no single operation can possibly satisfy the requirements of all ruptured disks. The operation must be tailored to the size and location of the offending fragment. Thus, the neurologic examination and precise imaging remain essential factors in preoperative evaluation. The possible association of facet arthropathy with spinal stenosis must also be kept in mind (Fig. 4).

The common surgical presentation is rupturing of the disk posterolaterally at the level of the disk space. Depending on its size, the herniation compresses the nerve root crossing the space and may distort the lateral cauda equina as well.

**Surgical Technique.** Surgical exposure of the compressed nerve root begins with a precise midline skin incision and paramedian incision in the deep fascia, permitting dissection of the paravertebral muscle mass with a fine electrocautery tip to expose two and one-half hemilaminas or two hemilaminas and the sacrum on the involved side. The length of the incision is adequate with the aid of paralytic agents given during anesthesia for atraumatic retraction of the muscle to the lateral facet margin. Hemilaminectomy above the superior margin of ligamentum flavum (L4 at the L4–5 disk, L5 at the L5–S1 disk) ensures that bony removal will be cephalad to the disk space. After full excision of the exposed ligamentum flavum, bone with the lateral shelf of ligamentum is taken away just beyond the lateral margin of dura by working from cephalad to caudad.

In this situation the nerve root is tightly compressed against the upper margin of the lamina below and the superior portion of facet. To gain its release, this bone is removed by working medially to laterally and starting where the dura is more slack. Enough bone should be taken away to ensure exposure of the compressed nerve root and lateral cauda equina cephalad, caudad, and lateral to the underlying fragment of disk. Adequate room is available for dissection and control of epidural veins. When the fragment is extruded, it comes immediately to view with slight retraction of the nerve root and can be dissected free and drawn laterally with traction. Such a lateral approach eliminates the need for nerve root retractors. The nerve root is easily displaced medially when so released (Fig. 5). When the fragment is not extruded but gives the appearance of a large, smooth pearl, the overlying distended longitudinal ligament is incised to allow extrusion of the fragment.

When not completely extruded, the fragment of herniated disk usually breaks off at its incarcerated stem and requires an incision in the anulus lateral to the tear for removal. The disk space is then entered for evacuation of as much of the remaining degenerated disk material as possible. It would seem unnecessary to warn about anterior penetration, yet reports of vascular, ureteral, and intestinal injury still appear. The end of the extracting instrument should be in view most of the time and in any event should not be introduced into the disk space beyond 2.5 cm. The recurrence rate of 6%–7% has not, in my experience, been altered by radical excision of intradiskal material and cartilaginous plate compared with removing as completely as possible the degenerated and loose fragments of disk that are accessible.

Removal of bone and surgical exposure must be modified when dealing with the situations in which the ruptured fragment of disk occupies a median or paramedian location, has extended above or below the disk space through a marginal tear, is large enough to cause serious compression of the cauda equina, or is associated with appreciable spinal stenosis caused by facet arthropathy. These situations occur in a number of patients who truly require surgical intervention for a ruptured disk. For example, about 15% of ruptured fragments occupy a median or paramedian position in the spinal canal (5). Also, the rare situation can occur in which a large ruptured paramedian disk fragment extends toward one side and produces contralateral radicular pain (Fig. 6).

Each of these situations requires sufficient decompressive exposure to release the neural elements and excise the offending fragment of disk completely. Thus, it is not at all uncommon for a fragment of extruded disk literally to straddle the L5 nerve root intraspinally at the entrance into its foramen and even into its neural canal. Whether the fragment has erupted from a superior marginal tear at L5–S1 or an inferior marginal tear at L4–5, the appropriate surgical approach requires full hemilaminectomy of the L5 segment. This approach permits adequate exposure and intraoperative dissection of the lesion so that the surgeon can decide which disk space to enter for retrieval of further degenerative disk material by seeking out the marginal tear—a sometimes impossible task without adequate exposure.

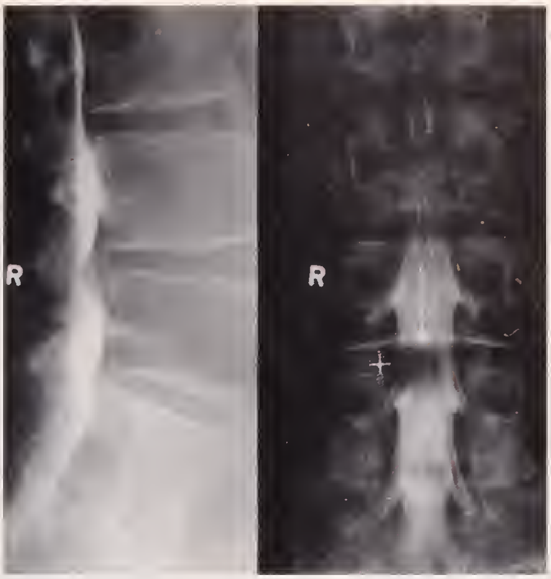


Fig. 6. Lumbar myelograms showing right paramedian defect at L4–L5 with compression of cauda equina extending to left of midline.

The large median or paramedian disk fragment demands lumbar laminectomy. The lesion is often massive and commonly produces complete blockage or obstruction of most of the spinal canal as seen on myelography, magnetic resonance imaging (MRI), or computed tomography (CT). Because most of these disk herniations produce unilateral pain and neurologic signs, laminectomy should be carried well laterally on the side of the symptoms and signs and leave the facet intact on the opposite side (Fig. 7). However, the operation should begin at the lower margin of the lamina below (lower margin of L5 for disk rupture at L4–5) and include most, if not all, of the lamina above, just as would be indicated in the presence of an intraspinal tumor of the same size.

**Easier Recovery with Less Invasive Procedures?** The contention that so-called less invasive procedures through minimal exposure result in easier recovery for the patient must be chal-

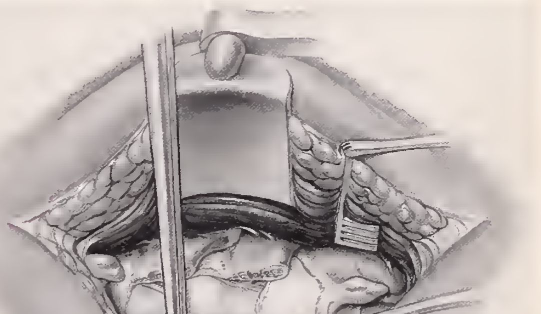


Fig. 7. Laminectomy exposure for large midline rupture of disk. Lateral bone removal on one side permits minimal retraction after release of cauda equina. Note mass displacing cauda equina. *Reprinted* with permission from Fager CA. Ruptured median and paramedian lumbar disk: a review of 243 cases. Surg Neurol 1985;23:317.

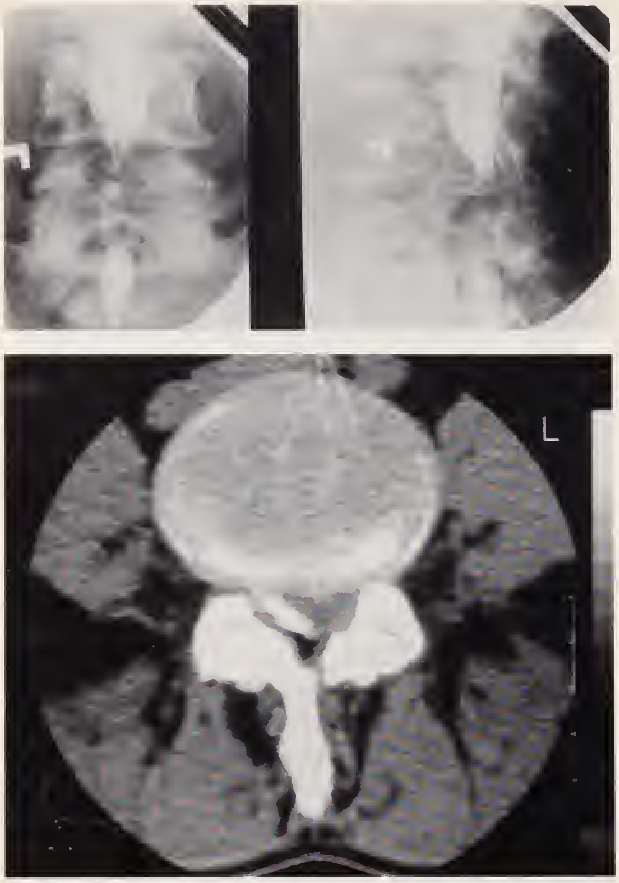


Fig. 8. Conventional myelogram and lumbar computed-tomographic scan after myelography showing median-paramedian rupture of disk.

lenged. With careful atraumatic dissection and adequate release of neural elements, the patient with a ruptured disk who requires laminectomy is hospitalized for 24–48 hr, usually gains immediate relief of pain, becomes active within days, and is back to work in periods varying from 1 to 6 weeks, depending on the type of work (Fig. 8).

**Multiple-Level Disk Excision.** The other controversy that should be addressed is the concept of multiple-level disk excision, either by open surgery or by intradiskal procedures. I have never encountered a patient with simultaneous symptomatic rupture of a disk at more than one level. But, as during the active use of Chymopapain, many patients now undergo disk procedures at multiple levels by the technique known as percutaneous diskectomy. Many of these multilevel operations have come about as a result of myelography, CT, or MRI findings of disk prominence, often reported as "bulging," when this phenomenon may not be symptomatic at all. Some prominence of the intervertebral disk is not abnormal at L4–5, and prominence of ligamentous, fibrous, and, at times, calcific tissue is obviously evident in degenerative narrowing of the disk space. The surgeon who undertakes lumbar spine surgery must be prepared to evaluate all such diagnostic procedures in relation to symptoms and signs and to assess the necessity and possible value of surgery.

## References

1. Fager CA. Atlas of spinal surgery. Philadelphia: Lea & Febiger, 1989:7.
2. Fager CA. The neurosurgical management of lumbar spine disease. New Dev Med 1988; 3:5–31.
3. Fager CA. Lumbar microdiscectomy: a contrary opinion. Clin Neurosurg 1986; 33:419–456.
4. Fager CA. Microsurgical intervention for lumbar disc disease: a critique. In: Williams RW, McCullough JA, Young PH, eds. Microsurgery of the lumbar spine. Rockville, MD: Aspen, 1990:249–262.
5. Fager CA. Ruptured median and paramedian lumbar disk: a review of 243 cases. Surg Neurol 1985; 23:309–323.

# Microdiskectomy—Myth, Mania, or Milestone? An 18-Year Surgical Adventure

Robert W. Williams, M.D.

## Abstract

The term "microdiskectomy" in the medical literature defines absolutely no surgical parameters or specific techniques for the treatment of the herniated nucleus pulposus. I attempt to clarify frequently used slang and to design a logical mental and microsurgical approach to the problem of lumbar disk herniations. The rigid surgical technique of microlumbar diskectomy is closely examined.

During the past eighteen years, the surgical microscope has been utilized by this author as a technical assistant during all lumbar operations (1–7). Not only has the clinical experience been rewarding, but the viewing of surgical pathology has been extensive. A great deal has been learned about lumbar disk disease, its often associated bony and ligamentous pathologies, and the iatrogenic morbidities of the patient whose lumbar surgery has failed (4). Physiologic and microsurgical concepts that have evolved during this personal lumbar adventure would seem worth sharing.

The word "microdiskectomy" is an eye-catcher and has been purposely used in the title of this manuscript to entice the reader. Clarification of the term and what it implies in the world medical literature, however, might prove to be a revelation. Although it grieves me to make such a disclosure, the word "microdiskectomy" portrays only the presence of a magnifying instrument in front of the surgeon's face during operations performed on the lumbar intervertebral disk. Additionally, all related terms, such as "microdisk," "microsurgical lumbar diskectomy," and "lumbar microsurgical diskectomy," depict a similar event. Collectively, such metaphors do not define any surgical parameters for the treatment of lumbar disk disease. Their only common denominator is the inferred use of a surgical microscope by the operator. The reader should clearly understand such terminology, since the concept the terms describe places no restraints on the surgeon or on the degree to which the lumbar anatomy gets "surgerized." Microlumbar diskectomy, however, is an entirely different story—namely, a rigid surgical technique (1, 2, 5–7).

As a responsible physician, I will not venture beyond certain lumbar surgical boundaries. Such self-imposed discipline has resulted from attempts to help numerous individuals with failed-back-surgery syndrome. If one observes enough of the misery sustained by these human beings, with their narcotic addition, sexual impotence, and economic devastation, frequently reoccuring clinical and radiographic patterns become evident. First of all, the chronic pain and neurologic deficits from which these individuals continue to suffer rarely resembles the original clinical syndrome for which an initial lumbar operation was prescribed. Second, a review of their preoperative and postoperative radiographic studies consistently reveals narrowing of one or more intervertebral disk spaces and/or surgical injury to one or more facet joints. Such realities dictate that these particular structures must be absolutely protected in the operating room.

---

From the Department of Neurosurgery, University of Nevada School of Medicine, Las Vegas, NV. Address reprint requests to Robert W. Williams, M.D., 3201 South Maryland Pkwy. #615, Las Vegas, NV 89109.

A technology has existed for a number of years now that allows any motivated surgeon to design vastly less traumatic and far more accurate lumbar operations. That technology is, of course, the surgical microscope, with its high-intensity light source. Utilizing this instrument, surgical nerve root injuries, blood transfusions, cerebral spinal fluid leaks, and significant wound infections can be virtually eliminated. Compressive neural pathologies, such as the herniated lumbar disk and the bony and ligamentous hypertrophies often associated with it, can be easily modified to resolve clinical sciatica while preserving the stability of the lumbar spine (6). Finally, postoperative and long-term surgical morbidities can be reduced to an absolute minimum (3). Unfortunately, the mastering of lumbar microsurgical technique and instrumentation cannot be acquired through casual encounter. Even if competence in the technology is achieved, a poor diagnostician will always produce poor surgical results.

## Anatomy of the Lumbar Spine

The lumbar anatomy must be appreciated by all physicians as an incredible piece of biological machinery. Although dysfunctions do arise from years of environmental stress and trauma, the lumbar spine is still a far better piece of engineering than anything humans have created for a similar purpose. A basic understanding of functional structure is paramount, however, if one is to attempt to design consistently successful surgical techniques for the resolution of lumbar neural compressive pathology.

**Lumbar Spine Design and Function.** The lumbar spine supports approximately 70% of the body's weight, except when the body is lying flat. Add the mechanical stresses of flexion, extension, and rotation demanded by modern living and it becomes easier to appreciate the segmental design of the lumbar anatomy that has apparently proved necessary to successfully accomplish the task. Additionally, the delicate cauda equina is housed within this segmental bony tube. To surgically collapse the intervertebral spaces or destabilize the alignment of the basic lumbar blueprint is to render the organism disabled.

**Stability.** Nature's ability to maintain an appropriate volume for lumbar neurologic tissues over the years seems to be the key to avoiding spinal surgery. For each individual human being, there is a time-honored race between the genetic, endocrine, degenerative, and mechanical stress factors that determine the adequacy of lumba epidural space at any particular moment. Sinc these factors cannot always be controlled, biological degeneration over time can be expected t take its toll on the lumbar anatomy and it treasured cauda equina.

Spinal tissues such as bone, ligament, an intervertebral disks are most certainly provide with pain receptors that alert the organism t physiologic dysfunctions. In my experience, however, such an alert can be classified as a mere annoyance compared to the miseries produced b neural compressive pathology. In the vast majority of cases, sciatic pain leads the patient to accep a lumbar surgical attempt for relief, rather tha pursuing a more conservative approach such a modifications in body weight and individual lifestyle.

Stability of segmental design maintains adequacy for lumbar neurologic tissues over th years. Hand any structural engineer an anatomi model for analysis, and the answer will always b the same: the intervertebral disks, the face joints, and their supporting ligaments stabiliz the volume of the lumbar epidural space. Whe these structures are themselves involved in neural compression, they can only be remodele within those limits that successfully preserv lumbar alignment and segmental spacing.

If these engineering principles are not respected in the operating room, the patient wil have acquired a new disease. I choose to term thi iatrogenic affliction "failed lumbar surgeon' syndrome." Additionally, middle-aged wome who are physically inactive, overweight, an have some degree of osteoporosis secondary t premature ovarian failure are candidates fo midlife lumbar instability at the L4–5 level When surgically treating lumbar disk disease i these individuals, even more extreme cautio must be exercised.

**The Lumbar Disk.** The intervertebra spacers, both the anulus fibrosus and the nucleu pulposus, are unique structures; their fibroelasti tissue is avascular and their nutrition is provide by some type of diffusion process through the adjacent vertebral cartilagenous endplates. Additionally, the volume of disk tissue seems to b within critical limits for the maintenance of adequate intervertebral spacing and segmenta alignment.

Disk tissue does not regenerate, nor does it anular retaining wall heal by primary intention To create a permanent window in the anulus fibrosus with a scalpel and to macerate the nucleu pulposus and cartilagenous plates by curettement during the reduction of a simple disk hernia is to virtually destroy the nutrition and future competence of the structure. In this day and age of human tissue salvage techniques, how could anyone attempt to justify such surgical behavior?

### Surgical Engineering for the Lumbar Spine

If one can accept that the integrity of the facet joints and the volume of intervertebral tissue are crucial to the physiology of the lumbar spine and its enclosed neural tissues, a challenge arises: to engineer methods for the protection of these structures during the surgical treatment of compressive neural pathology. For me personally, the two greatest adjuncts for such an endeavor have been the surgical microscope and magnetic resonance imaging (MRI). With such technology at one's disposal, it becomes possible to practice "designer lumbar surgery" for each and every patient.

There are really only two categories of patients for lumbar surgery: those who have never had an operation (virgin anatomy), and those who have previously been operated on but again develop sciatica. The latter group includes patients who experienced temporary surgical relief from initial radicular pain and those for whom operative intervention proved to be a total failure.

**Patients with Postsurgical Sciatic Pain.** I term any patient previously operated on who is again symptomatic with sciatica a patient with "postsurgical sciatic pain" (PSP) (4, 6, 7). Reexploration of this patient with the technical assistance of a surgical microscope can provide many interesting observations on the tissue morbidities of standard lumbar operations. For instance, the smaller the surgical wound, the less the possibility of incisional hyperesthesia and kinetic muscle spasms. Wherever the ligamentum flavum has been sacrificed, the tissue void will be filled with fibrous scar. Rongeuring of bone stimulates osteoblastic activity and osteophyte formation, which can subsequently lead to recurrent neural compression. Sacrifice of epidural fat tissue cultivates perineural fibrosis in the lateral recesses. Scalpel incisions in the anulus fibrosus heal only by secondary intention. Disk reherniations are encouraged by this technique, especially when curettement has been additionally applied to the intervertebral tissue. Finally, the use of electrocautery in the epidural canal generates dense scar tissue reactions. If one can accomplish the surgical relief of sciatica while avoiding tissue morbidities, the welfare of the lumbar disk patient will certainly be maximized.

Designer surgery for the lumbar patient is really quite simple if the surgeon obeys the following commandments:

- Utilize a radiographically localized minimal surgical exposure
- Avoid the surgical disturbance of all deep lumbar tissues except for those contributing to neural compression
- Minimize any sacrifice of ligamentum flavum and protect the epidural fat tissue
- Eliminate the use of electrocautery instruments in the lumbar spinal canal
- Once a disk hernia is reached, dilate an opening in the anulus fibrosus and decompress the herniation without curetting the fibroelastic intervertebral tissue (2, 4–7)
- Most important: avoid surgical injury to the facet joints and preserve a maximal volume of healthy disk tissue to maintain lumbar alignment and stability.

**Lumbar Clinical Skills.** Deficiencies in clinical diagnostic skills are only magnified in the operating room. Medical students, residents, and even fully trained lumbar surgeons can hardly wait for the anesthesia to begin. In my experience, the vast majority sorely lack basic lumbar diagnostic expertise. This deficiency may be the greatest single reason why lumbar surgical results are so marginal. How is it possible to engineer a successful nondestructive operation if the surgeon does not know why to operate, when to operate, where to operate, or what can be accomplished on arrival? In my opinion, there is a major academic void in the clinical expertise of most lumbar surgeons.

**Spinal Instrumentation.** The arena of lumbar spinal surgery seems somewhat to resemble a washing machine—the dirty laundry just keeps coming around. I feel sorry for those expert and dedicated lumbar-instability surgeons who are witnessing a national bloodbath, now that their hardware has been released for indiscriminate use by indiscriminate surgeons on the general public. In some hospitals, the indications for internal spinal fixation are (a) low back pain, (b) a herniated disk, or (c) anyone with a previous lumbar operation. Those indications encompass virtually the entire adult human population. Additionally, the placement of posterior internal fixation hardware may be performed one week prior to anterior lumbar interbody fusion.

The unremitting surgical morbidity on these patients has thus far proved to be incredible and will eventually be equaled only by the crippling financial morbidity to our compensation and health care system. Conservative microsurgical operations can successfully treat the same lumbar conditions while leaving behind an absolute minimum of surgical morbidity. Conservative lumbar microsurgery, however, is not nearly as remunerative.

**Lumbar Radiology.** Proficiency in correlation between clinical and radiographic examinations is mandatory in order to guarantee lumbar surgical success. Plain films of the spine with flexion and extension views are of value for visualizing congenital anomalies, suggesting the demineralization of bone by osteoporosis, and grossly evaluating degenerative diseases and lumbar instability. When the disk spaces are collapsing and the facets joints are enlarging from osteoarthritis, back pain can be expected to continue regardless of surgical ingenuity.

The CT scan has been of great value for visualizing certain types of lumbar neural compressive pathology. Sagittal and axial views should be obtained at all lumbar levels. The dose of radiation received by the patient during CT scanning is considerable, and the diagnostic quality of the study can be enhanced when combined with myelography. If MRI cannot be obtained, CT scanning with or without myelogram is the lumbar diagnostic study of choice.

MRI is state of the art (6). Spine surgeons should spend time with a radiologist knowledgeable in imaging technique. The full gamut of lumbar pathology can be assessed and the patient does not receive any radiation. With the surgical microscope as an adjunct, MRI has made designer lumbar surgery a reality.

Both CT and MRI scans are obtained with the patient lying flat; the dynamics of the intervertebral tissue may be poorly assessed. In those patients with a pure weight-bearing sciatica, myelography in the upright position may visualize a dynamic intervertebral disk herniation. Myelography otherwise has been of little value in my practice since the advent of MRI.

The responsibility for a lumbar surgical tragedy lies with the surgeon and not the radiologist. Before proceeding to the operating room, critically recorrelate the patient's clinical symptoms and objective findings with the radiographic studies. Always seek consultation with someone more experienced if in doubt. Failure of lumbar surgery is not the end result of a naturally occurring disease process, but rather belongs directly to the "failed lumbar surgeon."

## Rationale for Designer Surgery

An organized thought process is the secret to lumbar surgical success. If the wheat cannot be separated from the chaff, the patient does not stand a chance.

Chronic low back pain arises primarily from unremitting lumbar muscle spasm, degenerative bone diseases such as osteoarthritis and osteoporosis, and occasionally from isolated degenerative disk disease. Most often, the etiology will prove to be some combination of these conditions. A thorough clinical, medical, and radiographic evaluation will always resolve any quandary. The most effective treatment program will encompass modifications in lifestyle and reduction of excessive bodyweight. Thus far, no lumbar operation has proved successful for the primary treatment of low back pain caused by degenerative disease.

Manual labor and excessive body weight are killers of the lumbar spine. Additionally, the ravages of indiscriminate hysterectomy in young women are beginning to add osteoporosis to the list. There are no lumbar operations that will consistently prove worthwhile for more than temporary relief of sciatica in the presence of continued heavy labor or obesity. Without commitment from the patient, long-term lumbar surgical successes will be few and far between.

The vast majority of conditions amenable to lumbar surgery will be ongoing, progressive, and degenerative. Trauma may precipitate the onset of sciatica, but in the vast majority of cases the stage will be set before the play takes place. Learn to work within the segmental design of the lumbar anatomy while preserving its stability and alignment. Even in dilapidated condition, the surgically untouched lumbar spine will serve the individual far better than any reconstructive spinal operation being offered to the public today. When surgeons learn to design nondestructive rather than destructive lumbar operations, the end results will be optimal.

There are really only two categories of "virgin anatomy" lumbar neural compressive pathology. One is the herniated disk and the second is any variety of bony and ligamentous hypertrophy which irritates the nervous system by spinal stenosis. Of all the patients in my practice with first-time sciatica since the advent of the scanners, approximately 20% had lumbar disk herniations in pure form, 65% had disk herniations associated with stenosis, and 15% had stenosis alone. In other words, 85% of my patients with initial sciatica had protrusions or displacement of intervertebral tissue irritating the nervous system. If one could develop a reasonable scientific approach for the successful long-term modification of a disk hernia, the often associated but relatively nondynamic stenosis pathologies would be a simple matter.

Since the stability and alignment of the lumbar spine appears to be greatly influenced by appropriate intervertebral spacing, salvage of a maximal volume of healthy disk tissue would seem paramount when treating a disk hernia. Additionally, if the anulus or outer retaining wall of the disk could be penetrated in such a way as to best preserve its future competence, perhaps the surgically reduced disk herniation would demonstrate minimal spontaneous recurrence rate. With technology available which allows the herniated disk to be reached without significantly disturbing the normal intervening lumbar tissues, the best possible surgical design could be achieved—the surgical microscope and microlumbar diskectomy.

## Microlumbar Diskectomy

Microlumbar diskectomy (MLD) is a totally microsurgical technique designed specifically for the treatment of initial lumbar disk herniations in the absence of any other spinal canal pathology. A midline radiographically controlled minimal surgical exposure is utilized. The surgeon stands on the side of the disk hernia with the patient in the prone position. Spinal frames are not used but rather a vacant space is created under the abdomen by placement of a collapsible urologic irrigation bladder. Following scalpel incision to the fascia, an electrocutting knife separates the paravertebral fascia from the spinous processes. Finger dissection is used to displace the musculature laterally toward the surgeon and an MLD retractor is placed to expose the interlaminar space.

Without any degree of laminectomy or facet trauma, the lateral ligamentum flavum is removed with microsurgical rongeurs to expose the perineural tissues. Epidural fat tissue is preserved and the neural elements displaced medially with an MLD suction retractor. Electrical instruments are never allowed in the spinal canal. The anulus of the herniated disk is dilated with a blunt instrument, and micronerve hooks and MLD disk forceps are used to decompress the disk hernia. All healthy disk tissue is preserved, and the disk space is never curetted.

Wound closure includes only the fascia, subcutaneous tissue, and skin with absorbable sutures. The incision is approximately one inch in length, and the operating time about 37 min.

If microsurgical technique is rigidly followed, there should be an absolute minimum of surgical injury to the normal lumbar anatomy. Obviously, if one is treating disk herniations in patients with postsurgical sciatic pain (PSP) or spinal stenosis, only that portion of the technique directly related to dilation of the anulus fibrosus and salvage of all healthy intervertebral tissue is applicable. MLD technique is only beneficial in the latter two groups of patients, once the disk hernia is reached. All additional pathologies commonly found in patients with PSP and spinal stenosis cannot be successfully treated through such a limited surgical exposure.

With the rare exception of those patients whose lumbar disks genetically turn to mush over the years, the long-term outlook for any patient with a surgically treated herniated disk should be directly related to the health and volume of the residual intervertebral tissue and the postsurgical competence of the anulus fibrosus. MLD technique gives the operated-on lumbar disk the best chance of functional survival.

**Postoperative Management.** The postsurgical lumbar disk patient should be managed with vigor. Although compressive neural pathology may appear to have been successfully relieved during the static nonweight-bearing position at operation, the situation will change immediately when the patient assumes an upright posture. Competent healing of the anulus fibrosus and the sustained health of any residual intervertebral tissue will ultimately determine the long-term prospects for each patient. I am convinced that the individual's behavior during the first six months following surgery will account for 90% of any long-term operative success.

**Postoperative weeks 1 through 4:** No car riding in the sitting position; no housework, especially vacuuming; no water beds; and no lifting or bending. Stairs may be climbed only once a day. Walking is allowed as tolerated, but two hours of flat bed rest are required twice a day. Sexual activity is encouraged. Sitting straight-leg raises are performed mornings and evenings, ten with each leg. No other lumbar activity is allowed. Appropriate dieting is repeatedly stressed. Wound inspection in the office occurs seven days after hospital discharge, and the first follow-up visit is in 4 weeks.

**Postoperative weeks 5 through 8:** Car riding is allowed in the sitting position during week five, and driving is permitted thereafter, but no stick-shift cars, no sports cars, and no pickup trucks. Light housework can commence, but no bending, lifting, or vacuuming. Dieting should be strict as indicated. Daily swimming is encouraged with breast and side stroke, but no hot water Jacuzzi jets to the lumbar region. Discussions relating to vocational rehabilitation are initiated with individuals previously employed in heavy manual labor. Frequent sexual activity is encouraged, but no sexual gymnastics. Tennis players may volley the backboard, and golfers are permitted putting and short-iron practice. Sports such as basketball, baseball, horseback riding and motocycling inflict violent lumbar stress and are permanently discouraged. Patients not involved in occupations requiring heavy labor may return to work as tolerated after the sixth week. Sitting straight-leg raises should continue twice a day.

**Postoperative months 3 through 6:** Lifting is restricted to 10-lb weights. Light gardening and moderate housework are permitted. Vacuuming is still discouraged. Tennis buffs may play doubles, and 9-hole rounds of golf are allowed using only the irons. Swimming is to continue, but no diving. Normal body weight should be achieved. Vocational rehabilitation is encouraged to completion. Light workouts with Nautilus-type equipment and water aerobics are permitted.

**Postoperative months 7 through 12:** Patients are allowed to lift 25-lb weights. Water and snow skiing are permitted, as well as tennis singles and full golfing activities. Bowling and light games of racketball seem to be well tolerated. Sitting straight-leg raises twice a day should continue indefinitely. Activities such as jogging and free-weight training are permanently discouraged. After this time period, lifting injuries and chronic obesity will be the major factors associated with recurrent lumbar disc herniation.

**Surgical Results with Microlumbar Diskectomy Technique.** I have used the surgical technique of microlumbar diskectomy in the treatment of 1,017 patients during the past 18 years. The initial criteria are a herniated disk syndrome and the absence of any additional spinal canal pathology. In this series, 164 patients have required reoperation (Table 1); overall postsurgical results (3, 5–7) are summarized in Table 2.

For those 54 patients with recurrent disk herniations at the exact location where the surgery was performed, 90% reherniated within nine months of surgery, and 65% of all recurrences required a new lumbar injury before decompensating. A spontaneous reherniation rate of only 1.8% would seem to support the MLD technique of dilating the anulus rather than scalpel incision and not curetting the intervertebral tissue. New disk herniations totaled 4.1%, compared to 5.3% for recurrent herniations.

TABLE 1

*Reoperations by Site of Recurrence on 164 Patients after Microlumbar Diskectomy in 1017 Patients with Virgin Disks at Initial Surgery, 18 Years*

| | No. of cases | Reoperations % of 164 | Reoperation % of 1017 |
|---|---|---|---|
| Total recurrent HNPs | 54 | 33.0 | 5.3 |
| Spontaneous RE-HNPs | 18 | 11.0 | 1.8 |
| New-level HNPs | 29 | 17.7 | 2.9 |
| Opposite side, new HNP | 12 | 7.3 | 1.2 |
| Missed disks (1979) | 3 | 1.8 | 1.1 |
| Adhesion sciatica (1980) | 17 | 10.4 | 1.7 |
| Spinal stenosis (evolved) | 47 | 28.7 | 4.6 |
| TOTAL | 180 | | |

TABLE 2

*Postsurgical Results of Microlumbar Diskectomy on 1017 Virgin Disks, 18 Years*

| | No. of patients | % of 164 | % o 101 |
|---|---|---|---|
| MLD series | 1017 | | 100. |
| Re-op MLD series | 164 | 100.0 | |
| Surgical "cure" (1 MLD) | 853 | | 84. |
| Surgical "cure" (re-ops) | 149 | 91.0 | |
| Surgical failures | 15 | 9.1 | 1. |
| Total 18-year MLD "cure" | 1002 | | 98. |

Surgical "cure" in this series is defined as a patient who is econom cally productive if they so desire, physically comfortable without addi tive medication, and free from sciatic pain.

A wrong-level operation was performed on 3 occasions prior to 1979. The possibility of this surgical error has since been eliminated by obtaining a localizing lateral lumbar spine X-ray on all patients prior to skin incision. If any question remains on proper anatomic exposure during the operative procedure, a second x-ray is obtained. Recurrent sciatica secondary to adhesions at the site of the operation has not occurred since 1980.

The evolution of spinal stenosis is occurring at a rate of 4.6% in the 18-year series. When the process is associated with recurrent disk herniations, stenosis usually appears bilaterally in the recesses at the same level where the diskectomy was performed. The earliest occurrence was 3 years after surgery. In more recent cases, MRI has demonstrated dehydration and narrowing of the disk space.

**Other Applications for the Technique.** The technique is employed whenever possible in virgin-spinal-stenosis and PSP patients with disk herniations (6, 7). The quality of intervertebral tissue is generally much healthier in the virgin-stenosis patient, since the tissue has not previously been operated on. Observations in 558 PSP patients over the years has convinced me that scalpel incision in the anulus fibrosus and curettement of the disk tissue virtually destroy the future competence of the structure. It is extremely rare to see a recurrent disk herniation initially treated by microlumbar diskectomy with a gaping surgical wound in the anulus fibrosus and residual necrotic mush in the intervertebral space.

## Conclusion

If there is a myth in spinal surgery today, it is the term "microdiskectomy" and related medical nomenclature. Such lumbar microsurgical slang is worthless without technical definition to anyone except the surgeon's financial manager.

If there is a mania in spinal surgery today, it is the indiscriminate use of reconstructive lumbar operations. As any structural engineer could have predicted, the once-accepted standard lumbar surgical concepts and techniques of the past have finally proved to be destructive to the future competence of the lumbar anatomy. I predict that, with a minimum number of exceptions, the present mania for rods, plates, and screws will earmark an age of lumbar-surgical genocide.

Finally, if there could possibly be a recent milestone in spinal surgery, I believe one should critically evaluate the surgical concepts, technique, and long-term results of microlumbar diskectomy. Rigid microsurgical technique provides for minimal surgical insult to the patient, maximal preservation of future competence of the anulus fibrosus, and no future necessity for reconstructive lumbar operations. A spontaneous 1.8% disk reherniation rate in 1,017 patients over an 18-year period is hard to beat.

### References

1. Williams RW. Microlumbar discectomy. In: Surgical Techniques. Randolph, MA: Codman Shurtleff, 1977; 1–10.
2. Williams RW. Microlumbar discectomy: a conservative surgical approach to the virgin herniated disc. Spine 1978; 3(2):175–182.
3. Williams RW. Microlumbar discectomy: A 12 year statistical review. Spine 1986; 11(8):851–852.
4. Williams RW. Postsurgical lumbar scarring: a study of surgical morbidity. In: Lumbar Discectomy and Laminectomy. Rockville, MD: Aspen Publishers, 1987; 253–263.
5. Williams RW. Microlumbar discectomy: a surgical alternative for initial disc herniation. In: Lumbar spine surgery. Baltimore: Williams & Wilkins, 1988; 171–183.
6. Williams RW, McCulloch JA, Young PH. Microsurgery of the lumbar spine. Rockville, MD: Aspen Publishers, 1990.
7. Williams RW. Microsurgical techniques for the treatment of lumbar disc protrusions and associated pathology. In: Operative spinal surgery. London: Churchill Livingstone (in press).

# Lumbar Diskectomy under the Operating Microscope

Ved P. Sachdev, M.D.

I do, more or less, a "conventional" lumbar diskectomy using an operating microscope and have done so in over 700 patients in the last 17 years. I work and teach in a neurosurgical department strongly committed to the use of the operating microscope. The use of the operating microscope in lumbar disk surgery is an extension of the same principles that led to its introduction for other procedures—that is, a better stereoscopic view of a deep, small anatomical field approached through a relatively small incision under better lighting and needing less tissue dissection. I do not invent a new operation just by varying the amount of disk material that I take out or how I enlarge the opening in the anulus fibrosus. I do not name my procedure "designer surgery" because all surgeons have always designed their surgery to fit the needs of the patient, and I am frequently surprised to hear prominent neurosurgeons shun, even ridicule, the use of the operating microscope for lumbar disk surgery when they would not think of removing an acoustic tumor without one. I find this rigidity against the binocular microscope for lumbar disk operations unwarranted.

The term *microlumbar diskectomy* has crept into the medical literature in the past few years. The term is impossible to define precisely. The procedure is as individual as the number of microsurgeons who perform it. The only common factor seems to be the operating microscope itself. Some 20 or more years ago (1), the binocular microscope was brought into the operating room by a few pioneers in neurosurgery. In the beginning, courses and workshops were offered to those willing to learn. Now an operating microscope has become an essential tool in the neurosurgical operating theater.

I joined the Department of Neurosurgery at The Mount Sinai Hospital as a resident 21 years ago. Microsurgical technique had already been well established by Malis (1), and lumbar diskectomies were being performed using the operating microscope by some, if not all, members of our department. In 1974, I was coauthor of a paper (2) that referred to the use of the operating microscope at The Mount Sinai Hospital in most of 240 spinal operations, including 190 for herniated lumbar disks. One letter to the editor (3) about the article expressed the opinion that employing a microscope for lumbar disk surgery "smelled of gadgetry."

Our department has remained deeply committed to microneurosurgical techniques for the majority of cranial and spinal procedures. I have assisted almost 40 residents through our program, and most of these young neurosurgeons never saw a diskectomy performed without the use of an operating microscope. Many of the residents continued to employ the microscope for lumbar disk surgery in private practice, while others preferred magnification loupes for individual reasons not in any way connected to lack of faith in the rationale for microsurgical technique.

**Positioning the Patient.** Lumbar diskectomy under the microscope (Fig.), as practiced at our institution, has already been described in some detail (4, 5). Operating on a patient in the oblique or three-quarter prone position requires a period of adaptation for the surgeon; and, as for other kinds of surgery, there is a learning curve. The position of the patient—oblique, prone, or lateral—is far less important than the surgeon's knowledge of the anatomy involved.

Our choice of the oblique position for lumbar

---

From the Department of Neurosurgery, The Mount Sinai Hospital, New York, NY. Address reprint requests to Ved P. Sachdev, M.D., 1148 Fifth Avenue, New York, NY 10028.

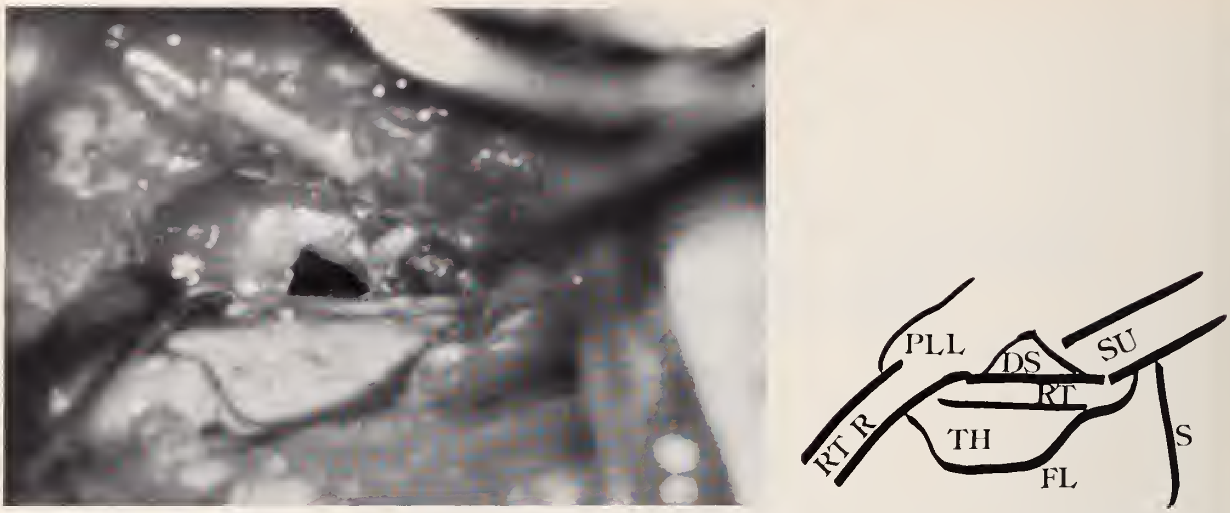


Fig. Lumbar diskectomy being performed under operating microscope. PLL, posterior longitudinal ligament; DS, disk space; SU, suction; RT R, root retractor; RT, root; TH, theca; FL, ligamentum flavum; S, skin.

disk surgery came after a few serious problems with the prone or the knee-chest position. A modified knee-chest or "Mohammedan-prayer" position (2) was standard for many years at The Mount Sinai Hospital until several instances of acute tubular necrosis were recorded. The acquisition of new operating tables did not permit the same type of positioning. As a compromise, the oblique position was easy to achieve even in relatively overweight patients and did not hinder respiratory excursions or cause pressure points. The intraspinal venous pressure was not generally affected, and any unusual increase in pressure could be reduced by lowering the head of the operating table slightly. Our anesthesiologists have had no difficulties with the patient in the oblique position. In our operating room setup, the nurse and the draped microscope are in front of the patient and the surgeon and the assistant are at the back. Aside from minor changes in the type of objectives and lenses and the need for an arm rest, the layout is the same for sitting surgeons and standing surgeons. A team of well-trained microneurosurgical nurses is essential, and we have been fortunate enough to have credible people doing an incredible job.

**Length of Skin Incision.** The length of the skin incision for lumbar diskectomy, with or without an operating microscope, has been overemphasized. I make a skin incision that is about 1 in. in length, but I do not let my ego get in the way when a slightly bigger opening is necessary to do the job well.

**Amount and Method of Lamina Removed.** That principle is also applied to the amount of lamina removed and the method of removal. I consider the term *minilaminotomy* (6), based on the length of the skin incision or the amount of bone removed, unnecessary and the term *microlumbar diskectomy* (7), based on the amount of disk material removed or the size of the opening in the anulus fibrosus, misleading. Basically we are all doing a modification of a well-established procedure, namely, decompressing a nerve root compromised by extrusion of a part of the nucleus pulposus. Almost every neurosurgeon performs a laminotomy rather than a laminectomy for a single-level posterolaterally herniated disk, and almost everyone agrees that there is no need for a fusion in cases not associated with spinal instability.

Much has been made of how much of the lower border of the lamina above the level of herniation should be removed. Rongeurs have been condemned by Williams (7) as a cause of bony spur formation later on. The use of high-speed drills such as Midas Rex (Midas Rex Pneumatic Tools, Fort Worth, TX) or Anspach (Marli Medical Supplies, Hempstead, NY) is esthetically pleasing and fast in experienced hands, but also more dangerous than the use of a rongeur such as a 45° Kerrison punch. This factor becomes very meaningful when a senior surgeon is taking a resident through a case. I have always used a rongeur because of its easy availability, and I have never seen the type of secondary spur formation which, in Williams' opinion (7), contributes to encroachment on the spinal canal or causes root compression.

**Protecting the Nerve Root as Guide for Technique.** Some lumbar disk surgeons just reflect the ligamentum flavum as a flap and then replace it (8). I have found such technique unnecessarily restrictive. Rather than remove the entire ligamentum flavum from the interspace, I resect as much as I feel necessary for exposure. There should be no preconceived "conveyer-belt" surgery. The lateral thicker part of the ligamentum flavum, the so-called falciform ligament, is densely attached to the capsule of the apophyseal joint, and opening into the facet is frequently unavoidable. Removing the most medial part of the descending and ascending facet of the joint is often necessary and desirable for the safety of the nerve root under compression. The "prima donna" of the exposed anatomy is the compressed nerve root; early identification and protection and ultimately decompression should be the only objectives of surgery. Obtaining enough room, whether by removal of ligamentum flavum or removal of upper and lower laminae or removal of descending or ascending facets, is good surgery and good judgment. Resection of a significant part of the facet joint is frequently necessary for foraminal or far-lateral herniations. All the bone work can be readily achieved under the operating microscope.

**Method of Opening the Anulus.** Central disk extrusions can be easily managed under the microscope by bringing the sequestered fragment into the field of view with a down-cutting Epstein curette precisely placed underneath the nerve root and dural sac. Ideally, after going back into the intervertebral space, the disk fragment can be delivered through a laterally created window in the anulus fibrosus. I never hesitate to cut such a window with a no. 15 blade, and after excision of the extruded nucleus pulposus, the same knife is employed to enlarge the rupture in the anulus fibrosus. Entering the intervertebral space with straight or angled pituitary rongeurs allows removal of copious amounts of nucleus pulposus. A surprising amount of disk material is found when the initial fragment is desiccated. I do not believe that degenerated parts of the nucleus pulposus can be stopped from extruding through the defect in the anulus fibrosus through which the initial herniation occurred. Enlarging the opening in the anulus fibrosus by a blunt instrument such as a Penfield no. 4 instrument cannot conceivably make the rupture in the anulus act like a sphincter and prevent further extrusion (7). Analysis of the results reported by Williams (7), who appears to be the only protagonist of this method of opening the anulus, shows an almost 11% recurrence rate, with a true recurrence rate at the level of 3.5%—about the same recurrence rate as most surgeons are reporting.

**Recurrence Rate of 3%.** The early results of my series were reported in 1986 (4). An oral review was presented in 1985 at the combined New York Society of Neurosurgery/New England Neurosurgical Society meeting. Better than 94% good results are continuing after first surgery for single-level, posterolateral disk herniations. There is still a 3% recurrence rate at the same level confirmed by computer tomography or magnetic resonance imaging. No fatality or serious permanent neurologic deterioration has occurred. One patient who developed a postoperative hematoma and cauda equina compression required reexploration and then made an excellent recovery. One superficial infection was treated with appropriate antibiotics.

Selection of patients for lumbar diskectomy, clinical presentation, and noninvasive and invasive modes of treatment have been well discussed by many authors (7, 9–11). The surgical techniques and principles discussed are basically sound, though frequently with too much emphasis on "thou shall not . . ." commandments based on strong personal opinions. In my experience, the operating microscope has been a tremendous help, and I intend to continue using it—my way.

## References

1. Yasargil MG. Microsurgical operations for herniated lumbar disc. Adv Neurosurg 1977; 4:81.
2. Miller R, Tausk HC, Sachdev VP. Anesthesia for laminectomy in the "crouch" position. Anesth Rev 1974; 1:9–13.
3. Letters to the editor. Anesth Rev 1975; 2.
4. Sachdev VP. Microsurgical lumbar discectomy: a personal series of 300 patients with at least 1 year of follow up. Microsurgery 1986; 7:55–62.
5. Zappulla RA, Hollis PH. The Microsurgical Approach to the Herniated Lumbar Disc. In: Camins, O'Leary, eds. The Lumbar Spine. New York: Raven Press, 1987:143–148.
6. Savitz MH, Katz SS. Minilaminotomy, an alternative to microdiscectomy or standard laminectomy. Mt Sinai J Med 1986; 53:518–522.
7. Williams RW. Microlumbar discectomy: a surgical alternative for initial disc herniation. In: Lumbar spine surgery, Baltimore: Williams and Wilkins, 1983. 171–183.
8. Wilson DH, Harbaugh R. Microsurgical and standard removal of the protruded lumbar disc: a comparative study. Neurosurgery 1981; 8:422–427.
9. Finneson BE. In: Lumbar spine surgery. Baltimore: Williams and Wilkins, 1983; 78–84.
10. Macnab I, MuCulloch J. Disc degeneration with root irritation. In: Bachache, ed. Lumbar disc surgery, 2nd ed. Baltimore: Williams and Wilkins, 1986; 283–334.
11. Fager CA. The neurosurgical management of lumbar spine disease. New Devel Med 1988; 3:5–31.

# Automated Percutaneous Lumbar Diskectomy

Gary Onik, M.D.

## Abstract

Automated percutaneous lumbar diskectomy was introduced in 1985. Development of this technique was instigated by the general trend in medicine toward less invasive procedures. Technique, indications, patient selection, and a review of the current literature are presented. Cumulative data indicate that the procedure has the least potential to do harm compared to open operations for treating lumbar disk herniations.

Automated percutaneous lumbar diskectomy (APLD) is a relatively recent advance in the treatment of herniated lumbar disks. The technique involves placement under fluoroscopic guidance of a small, 2-mm automated suction probe into the affected disk under local anesthesia. A portion of the central nuclear material is then aspirated. The purpose of the procedure is to relieve the patient's symptoms by decreasing the intradiskal pressure; transmission of decreased pressure through the rent in the anulus to the herniation relieves the pressure on the affected nerve root. Development of this technique was instigated by the general trend in medicine toward less invasive procedures and the perception that the traditional surgical treatment for herniated lumbar disks was associated with variable results and significant morbidity.

In fact, the recent literature dealing with the morbidity of open diskectomy has emphasized the potential for possible complications. Stolke et al. (1), in a prospective study of experienced neurosurgeons, showed a complication rate of 13%, including one death, 3 nerve injuries, and an approximate 1% rate of diskitis in 481 cases. Ramirez and Thisted (2) reviewed 28,000 open diskectomies in a community setting and found that 1 in 64 patients had a major complication; 1 in 335 had a major neurologic complication; 1 in approximately 500 a cardiovascular complication; and 1 in 1,700 patients died. Sixty percent of these procedures were performed by neurosurgeons, and there was no difference in the complication rate between neurosurgeons and orthopedists.

Chymopapain was the first of the less invasive treatments for herniated lumbar disk. Chymopapain, unfortunately, did not meet the most important criterion for such a procedure, namely safety, since there were reported problems of anaphylaxis, severe back spasms in a significant portion of patients, disk space narrowing, and transverse myelitis with paraplegia. These serious complications greatly curtailed the use of chymopapain in the United States.

Percutaneous diskectomy, a mechanical decompression of the disk, is an attempt to gain the same results as chymopapain without the associated complications. Percutaneous diskectomy was introduced by Hijikata et al. (3) in 1975. Since that original paper, a number of other investigators (4–10) using similar manual techniques have published their results, with success rates of 72%–87%. All these hand methods of percutaneous diskectomy have in common a relatively large cannula (6 mm outer diameter or greater) for access to the disk space and removal of disk material by some form of grasping instru-

---

From the Departments of Radiology and Neurosurgery, Presbyterian-University Hospital, Pittsburgh, PA 15213. Address reprint requests to Gary Onik, M.D., Department of Radiology, Presbyterian-University Hospital, DeSoto and O'Hara Streets, Pittsburgh, PA 15213.

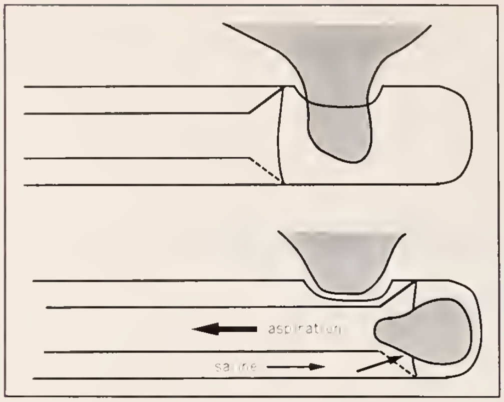


**Fig. 1.** Top Drawing of Nucleotome illustrating how suction is applied to central bore for aspiration of disk material into side port. Bottom Inner cutting sleeve pneumatically driven across side port guillotines disk material for suspension in saline and aspiration.

ment. Although the technique is efficacious, the most important criterion for a noninvasive procedure has not been met, namely significantly lower morbidity than open operations; two instances of major vascular injuries have been reported (4, 8). In addition, since the disk is entered multiple times, the rate of diskitis has been 1%, the same rate reported for open diskectomy (11, 12).

Automated percutaneous lumbar diskectomy, introduced in 1985 (13), takes advantage of a 2-mm reciprocating suction cutter that can cut separate pieces of disk material up to 200 times per minute. The Nucleotome aspiration probe works by drawing nucleus pulposus into its recessed side port. A flared inner cutter is pneumatically driven across the port and cuts off disk material, which is then suspended in saline and suctioned away (Fig. 1). Because anulus cannot be sucked into the side port, the Nucleotome cannot cut its way out of the disk; this feature adds to the safety of the procedure.

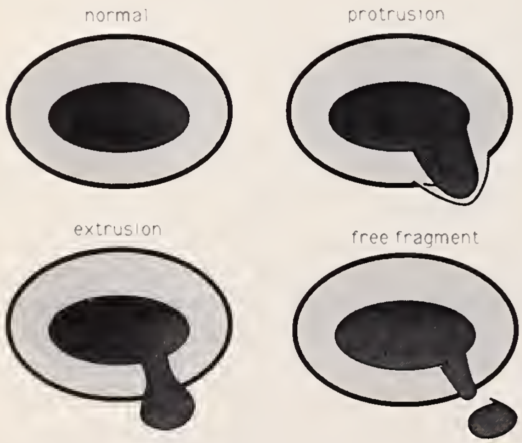


**Fig. 2.** Various stages of disk herniations.

By automating the process, the size of the instrumentation was reduced, thereby decreasing the chances for nerve injury, vessel damage, and soft tissue trauma. Once the instrument is placed into the disk, it remains there without being removed; the chances of contamination and diskitis are lowered to approximately one-fifth of the rate for open procedures or the hand method of percutaneous diskectomy (14).

There are now 21 studies on the procedure in print or in press (14–34) that include over 3,600 patients, proving that the concept is valid. Twelve of the reports show a success rate of 80% or greater, and seven of the studies show a success rate of between 70% and 80%. No major complications—specifically, no death, permanent nerve injury, or major vessel damage—have been reported, and the incidence of diskitis has been 0.2%.

Over 3,000 physicians have been trained in small-group workshops to perform the procedure, and over 40,000 cases have been completed worldwide (Henry Klyce, personal communication). In recent policy statements, the American Academy of Orthopedic Surgeons listed percutaneous diskectomy as an accepted procedure for the treatment of herniated lumbar disks. The National Blue Cross/Blue Shield Technology Assessment Committee determined the automated technique to be safe and efficacious in treating contained herniated lumbar disks and removed the procedure from investigational status.

## Patient Selection

As with all back surgery, good results depend on proper patient selection (ruled box).

**Clinical Criteria.** The procedure should be reserved for those patients whose pain is predominantly in the leg, secondary to a herniated disk, and not for patients with chronic or acute low back pain. Before the procedure is considered, all patients should have failed to respond to an adequate course of conservative care, at least 6 weeks of rest, analgesics, and physical therapy. In my practice, the patients have usually undergone three months of conservative treatment.

The procedure is not designed to treat patients who would otherwise be excluded from having surgery. Patients selected for the technique should have all the classic history and physical findings, as well as X-ray manifestations, associated with a herniated lumbar disk. From this clinical population, those patients

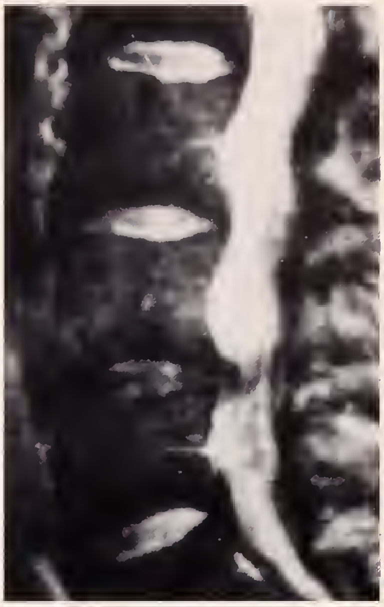
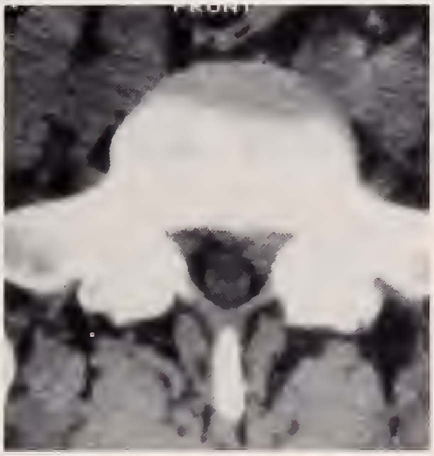
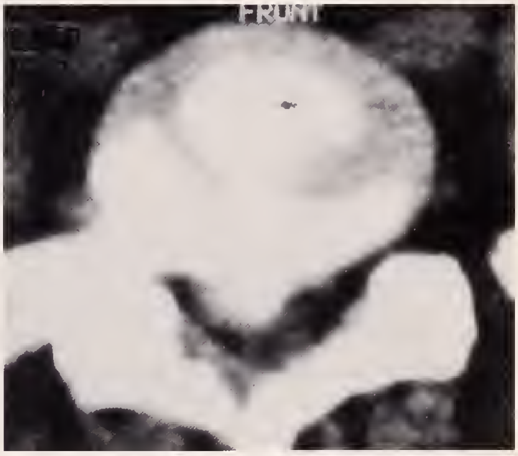


**Fig. 3. Right** Sagittal MRI scan. Note focal protrusion at L4–5 level. **Below right** CT diskogram showing focal protrusion with no evidence of free flow of contrast into epidural space. Patient was successfully treated by percutaneous diskectomy. **Fig. 4. Below left** Axial CT scan. Note free fragment in lateral recess.





whose herniations are still contained by the anulus fibrosus or posterior longitudinal ligament (protrusion) are singled out (Fig. 2). Performance in patients with extruded or free fragments would obviously lead to a large number of failures. The procedure is designed only for contained herniations.

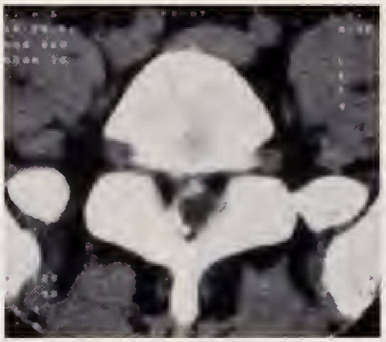
In a recent prospective surgical series (35), 20% of the patients who underwent an open procedure had contained disk herniations. Over 200,000 disk operations are performed each year in the United States; 40,000 of those patients might be spared a major surgical procedure by having the automated procedure instead.

**Radiologic Criteria.** The major burden in selecting patients for the procedure falls on radiologic imaging. Preoperative computed tomography (CT) or magnetic resonance imaging (MRI) scans should be performed to confirm the presence of a small-to-moderate focal bulge or herniation that is consistent with the patient's symptoms (Fig. 3) and to exclude a free disk fragment. Myelography alone is not adequate, since it does not exclude complicating abnormalities such as lateral recess stenosis, calcified disk herniations, or far-lateral herniations.

*Selection Criteria for Automated Percutaneous Lumbar Diskectomy*

**Clinical**
- Sciatica (leg pain > back pain)
- Physical findings that correlate with pain
- Failure of conservative therapy of at least 6 weeks

**Radiographic**
- Herniated disk contained by anulus or posterior longitudinal ligament
- Exclude:
  Free fragments
  Spinal stenosis
  Calcified disks



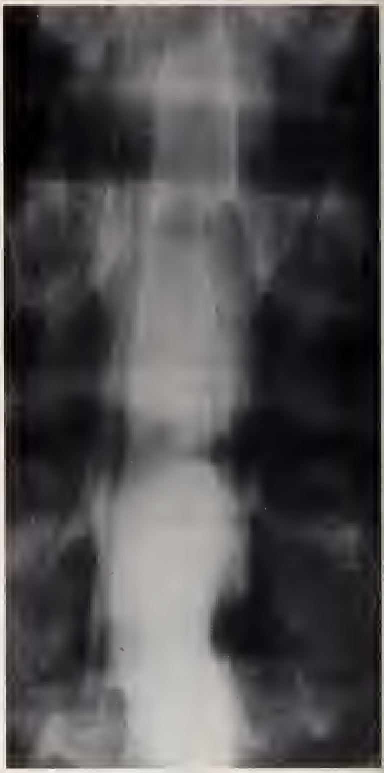


**Fig. 5. Left** CT diskogram demonstrating free flow of contrast material into epidural space and indicating noncontained disk. **Fig. 6. Right** Myelogram revealing focal disk herniation in a 15-year-old successfully treated by percutaneous diskectomy.

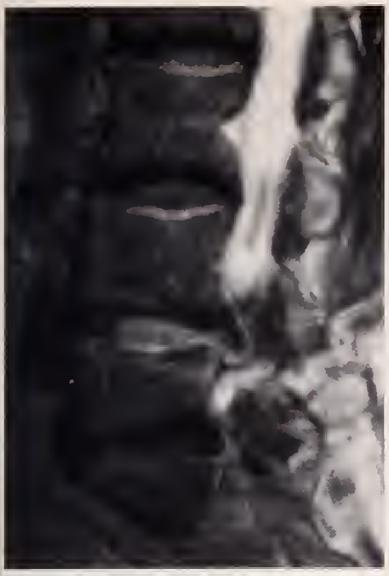
The size of the disk herniation is extremely important; in one report (36), 90% of patients with herniated disks that comprised 50% or more of the diameter of the thecal sac had extruded or free fragments. In addition, the shape of the herniation indicates whether the herniation is extruded or contained. Contained herniations tend to have smooth obtuse margins that blend at the base with the rest of the anulus. Irregular herniations are at sharp angles to the anular margin and probably have extruded fragments.

CT or MRI demonstration of a piece of disk material located above or below the disk space or filling a lateral recess is a contraindication because a migrated fragment would not be affected by central disk decompression (Fig. 4). When there is any doubt about extrusion through the posterior longitudinal ligament, diskography can investigate the integrity of the anulus and posterior longitudinal ligament. Free flow of contrast material into the epidural space indicates a complete tear of the anulus and posterior longitudinal ligament and a high likelihood of extrusion (Fig. 5). Demonstration of flow of contrast material through the anular tear and into the disk herniation indicates communication between the center of the disk and the herniated portion; extraction of disk material may still create a pressure difference between the disk herniation and the center of the disk.

When diskography does not demonstrate a herniation on CT or MRI, percutaneous diskectomy should not be undertaken. Edwards et al. (37) showed on CT diskography that herniations not stained by contrast injected into the center of the disk had only a 10% correlation with good results after administration of chymopapain, since the anular tear had closed and did not allow communication between the herniation and the nucleus. It stands to reason that the same criteria would be applicable to percutaneous diskectomy.

**Young Patients.** Symptomatic herniated disks that require surgery are unusual in young patients. In this patient population, physical therapy and conservative care usually resolve symptoms. When a young patient—aged 19 years or younger—does need surgery, the automated technique is the procedure of choice owing to its low morbidity as well as its excellent success rate. Ten patients in my practice aged 19 or younger have undergone the procedure successfully, with over 12 months of follow-up (Fig. 6). One reason for this high success rate may be the fluid quality of disks in young patients.

**Patients Who Have Previously Had Surgery.** Since patients who have had previous surgery and reherniation have a lower success rate with an open operation as well as increased risk of complications, the automated technique is the procedure of choice in this population. The pos-



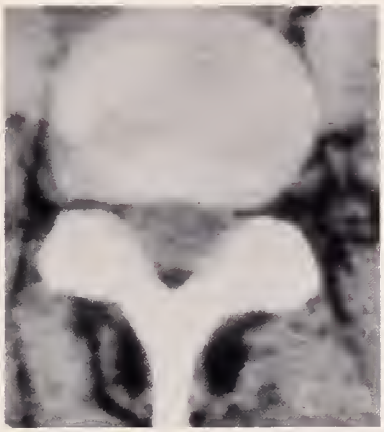


Fig. 7. **Left** MRI showing recurrent herniated disk with associated epidural scarring. **Fig. 8. Right** Axial CT scan showing far-lateral disk herniation. Both patients were successfully treated by percutaneous diskectomy.

terolateral approach avoids the spinal canal and is not affected by epidural fibrosis.

An acceptable success rate may be expected if certain criteria are met (Fig. 7). Forty-four patients who had previous surgery were selected to receive the procedures by Davis et al. (32) if the pain could be reproduced on a diskogram during diskography and contained herniation was demonstrated. Every patient had been symptom-free after the initial operation and then experienced a recurrence of symptoms. The 90% rate of success in 44 patients with a one-year follow-up may be attributable to a number of factors, including:

- This clinical population had already experienced open operation and wanted to avoid further major surgery.
- In addition, epidural fibrosis and scarring may have made disk fragment migration less frequent.

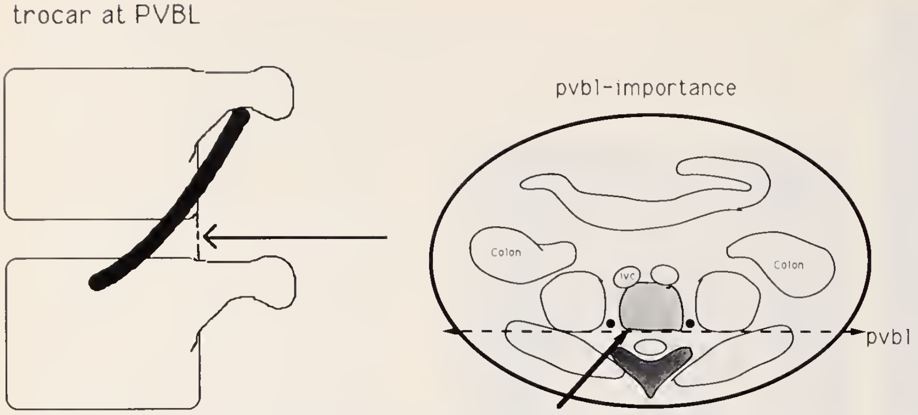
**Patients with Far-Lateral Herniation.** Far-lateral herniations (extraforaminal) can be a difficult problem. The posterior interlaminar open approach may necessitate partial or complete removal of the facet and possibly increase the chances for instability. Although the lateral approach may avoid the problem, this operation is unfamiliar to many neurosurgeons. Because of the position of the herniation, the instruments used in the automated procedure (38) can often traverse the abnormality (Fig. 8). Eleven patients with far-lateral herniations have undergone the automated procedure; 10 with more than one-year follow-up were successfully treated. In this group, the same criteria for fragment migration above and below the disk space should be carefully evaluated.

**Patients with Possible Diskitis.** The automated technique is the procedure of choice for possible diskitis, either primary or secondary to a previous operation. The low morbidity, the ability to obtain samples of tissue, and the decompression and debridement of the disk space make the procedure an excellent choice for this group of patients. In four patients who had negative needle biopsies, the automated procedure demonstrated positive culture results (39, 40).

## Technique

A complete, detailed description of the automated technique is beyond the scope of this paper; however, a number of technical points explain the low morbidity of the procedure. The path to the disk is posterolateral. The instruments pass just anterior to the superior articular facet, behind the nerve, in a trajectory toward the center of the disk, and avoid the spinal canal completely (Fig. 9). Certain radiographic landmarks prevent injury to vital structures during placement of the instruments. Staying posterior to a line that connects the posterior vertebral bodies, anterior to which lie the nerve, bowel, and great vessels, prevents injury to these vital structures (Fig. 9). When the instruments are against the anulus, they should be at or posterior to this line (Figs. 9, 10). On the anteroposterior view, an imaginary line connecting the medial border of the pedicles indicates the lateral border of the thecal sac. When the instruments are on the anulus and lateral to this line, they will not impinge on the thecal sac (Fig. 11).

The procedure must be performed with the patient under local anesthesia to allow the pa-



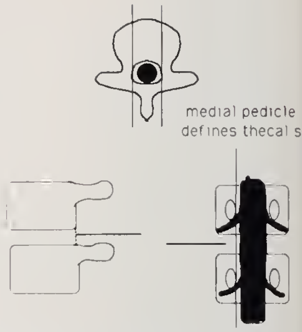


**Fig. 9.** Diagrams (left, middle) illustrating path instruments take during automated percutaneous lumbar diskectomy. Note importance of line connecting posterior vertebral bodies (pvbl); ivc, inferior vena cava. **Fig. 10.** Right Diagram demonstrating that when instruments are against anulus they are outside a line connecting medial border of pedicles.

tient to react to any nerve root impingement. If there is any complaint of radicular symptoms during the procedure, the instruments are redirected, and nerve root injury is prevented.

Localization of the instruments to the center of the disk is first carried out with an atraumatic 18-gauge trocar. Once the trocar is localized to the center of the disk, the larger, 2.8-mm cannula and dilator are brought down over the trocar to the disk. The dilator is removed, and the anulus is incised with a trephine. The trephine and trocar are removed, and the Nucleotome is then placed within the disk (Fig. 11). The procedure is monitored by observing the disk material aspirated into the tubing. When no further disk material is seen, the procedure is concluded, and the instruments are removed. The total time of aspiration is 10 to 20 minutes.

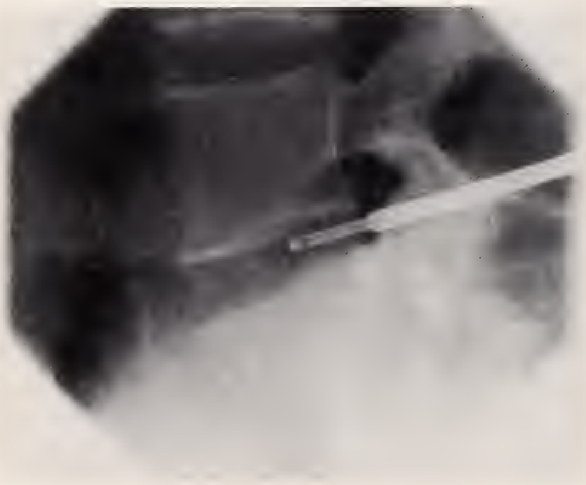


Fig. 11. Fluoroscopic lateral X-ray showing Nucleotome in disk space.

Positioning of the cannula on the anulus prevents the Nucleotome from injuring surrounding soft tissue structures if it is inadvertently withdrawn from the disk. In addition, the fact that the Nucleotome is not long enough to reach the opposite anulus and the fact that the Nucleotome cannot cut anulus keep the instrument safely within the disk space.

The approach to the L5–S1 disk space is essentially that of L4–5 and has previously been described in detail (41). The small size of the instruments and the introduction of a curved cannula, through which the flexible straight Nucleotome can be placed, allows entry into the L5–S1 disk in over 90% of cases (42).

## Reports of Results

**Prospective Multiinstitutional Studies.** The findings of a prospective multiinstitutional study to evaluate the automated procedure have been reported (15). Patients with classic sciatica who demonstrated a herniated lumbar disk consistent with their symptoms on CT or MRI were included. Patients were prospectively excluded if they had a history of previous lumbar surgery, previous chymopapain injection, worker's compensation claims, or other causes for their pain revealed by imaging modalities such as degenerative facet disease, spinal stenosis, or a free disk fragment. Criteria for success were stringent, the bottom line being that the patient had to be satisfied with the results of the procedure. Any patient who required a second procedure—another percutaneous diskectomy or an open operation—was considered a failure.

From November 1984 through May 1987, 506 automated procedures were performed by 18 different surgeons. Of the 506 cases, 11 were lost to follow-up and 327 met the prospective study criteria. However, in an additional 168 cases that did not meet the protocol, the procedure was knowingly done and reported in the study. The mean duration of preoperative conservative care was 11.5 months, allowing ample time for resolution of symptoms. In the 327 patients with a year or more of follow-up who were within the protocol, the success rate was 75.2%. In those patients who did not meet the protocol criteria, 49.4% of the procedures were successful. Of the 81 patients within the protocol considered to be failures, 44 underwent either laminectomy, microdiskectomy, or fusion; 19 underwent repeat percutaneous diskectomy; only 3/19 subsequently required an open operation. If the 16 patients who were successfully treated by a second automated percutaneous diskectomy were included, the success rate of the percutaneous procedure rises to 80%. Of the 44 patients who subsequently underwent an open procedure, 30 (68%) had free fragments not seen on preoperative imaging.

Most important, the findings in this study have been reproduced by other investigators. In another prospective multiinstitutional study conducted in Italy (15), 650 patients were treated with the automated procedure, the success rate was 72%, using the same selection and success criteria as the U.S. study. In a French prospective multiinstitutional study (16), 286 patients were treated with a success rate of 78%.

**Two Retrospective Series.** Only two retrospective series, of 17 and 27 patients with an overlap of 17 patients, reported a success rate less than 70% (22, 34); of note is the small number of patients in these studies, as well as the fact that patients with back pain alone were included.

**Low Morbidity.** Of critical importance is that the projected low morbidity for this procedure has been confirmed by numerous studies. There has been no reported incidence of permanent nerve damage, great vessel damage, or damage to the thecal sac and no mortality. A diskitis rate of 0.2%, three to five times lower than that of open diskectomy or the hand method of percutaneous diskectomy, has been documented.

Even if success rates of 95% reported for microdiskectomy are considered, the difference in efficacy between open surgery and automated percutaneous lumbar diskectomy is 15% to 20%. The advantages of the technique far outweigh the lower success rate primarily associated with missed free fragments of disk material. These advantages include:

- markedly decreased chance for nerve injury due to performance under local anesthesia
- no violation of the spinal canal with the posterolateral approach
- no creation of epidural fibrosis
- decreased chance of postoperative instability associated with removal of bone and ligaments
- low infection rate due to single entry of the disk space by automated probe
- outpatient procedure or same-day surgery
- return to work within two weeks (21).

Compared to chymopapain, the automated procedure has a low incidence of back spasm and virtually no risk of anaphylaxis or transverse myelitis. No bridges are burned. A patient who fails to gain relief of pain can always go on an open operation with no negative effects.

## Conclusion

Automated percutaneous lumbar diskectomy is both efficacious and extremely safe in treating patients with herniated lumbar disks contained by the anulus and posterior longitudinal ligament. The technique fulfills the most important criteria for a minimally invasive procedure: a reasonable success rate with minimal associated morbidity. The first goal of the physician is to do no harm. Compared to other available procedures for treating herniated lumbar disks, this procedure unequivocally has the least potential to do harm. Caution must be exercised in the careful selection of patients, and attention must be given to radiologic localization in order to ensure safe, effective results.

### References

1. Stolke D, Sollman W-P, Seifert V. Intra- and postoperative complications in lumbar disk surgery. Spine 1989; 14:56–58.
2. Ramirez LF, Thisted. Complications and demographic characteristics of patients undergoing lumbar diskectomy in community hospitals. Neurosurgery 1989; 25:226–231.
3. Hijikata S, Yamagishi M, Nakayama T, et al. Percutaneous diskectomy: a new treatment method for lumbar disk herniation. J Toden Hosp 1975; 5:5–13.
4. Hijikata S. Percutaneous nucleotomy: a new concept technique and 12 years' experience. Clin Orthop 1989; 238:9–23.
5. Hoppenfeld S. Percutaneous removal of herniated lumbar disks: 50 cases with ten-year follow-up periods. Clin Orthop 1989; 238:92–97.
6. Kambin P, Schaffer JL. Percutaneous lumbar diskectomy: review of 100 patients and current practice. Clin Orthop 1989; 238:24–34.
7. Monteiro A, Lefevre R, Pieters G, Wilmet E. Lateral decompression of a pathological disk in the treatment of lumbar pain and sciatica. Clin Orthop 1989; 238:56–63.

8. Schreiber A, Suezawa MD, Leu H. Does percutaneous nucleotomy with diskography replace conventional diskectomy? Eight years of experience and results in treatment of herniated lumbar disks. Clin Orthop 1989; 238:35–42.
9. Shepperd JAN, James SE, Leach AB. Percutaneous disk surgery. Clin Orthop 1989; 238:43–49.
10. Stern MN. Early experience with percutaneous lateral diskectomy. Clin Orthop 1989; 238:50–55.
11. Wilson DH, Harbaugh R. Lumbar diskectomy: a comparative study of microsurgical and standard technique. In: Hardy R, ed. Lumbar disk disease. New York: Raven Press, 1982:147–156.
12. El-Gindi S, Arefs AJ. Infection of intervertebral disks after operation. J Bone Joint Surg 1976; 58B:114–116.
13. Onik GM, Helms C, Hoaglund F, et al. Successful percutaneous lumbar diskectomy using a new aspiration probe: a case report. Am J Radiol 1985; 6:290–293.
14. Onik G, Mooney V, Maroon JC, et al. Automated percutaneous diskectomy: a prospective multi-institutional study. Neurosurgery 1990; 26(2):228–233.
15. Bocchi L, Ferrata P, Passarello F, et al. La neucleoaspirazione secondo Onik nel trattamento dell'ernia discale lombare analisi multicentria del primi risultate su oltre 650 trattamenti. Riv Neuroradiol 1989; 2:119–122.
16. Benazet JP, Roy-Camille R. Automated percutaneous lumbar diskectomy: results of French multicenter study of 287 cases. Rachis (in press).
17. Maroon JC, Allen RC. A retrospective study of 1,054 APLD cases: a twenty-month clinical follow-up at 35 US centers. J Neurol Orthop Med Surg 1989; 10(4):335–337.
18. Lesoin F, Autricque A, Clarisse J, Villette L, Jomin M. La Nucleotomie percutanee automatisee en pathologie discal lombaire. J Chir (Paris) 1989; 126:185–188.
19. Hammon W. Percutaneous lumbar nucleotomy. Neurosurgery 1989; 24:635.
20. Flynn LM, Morrow RE, Brannon J, et al. Arthroscopically assisted lateral percutaneous lumbar diskectomy using the Onik Nucleotome. J Neurol Orthop Med Surg 1988; 9:45–47.
21. Davis GW, Onik G. Clinical experience with automated percutaneous lumbar diskectomy. Clin Orthop 1989; 238:98–102.
22. Goldstein TB, Mink JH, Dawson EG. Early experience with automated percutaneous lumbar diskectomy in the treatment of lumbar disk herniations. Clin Orthop 1989; 238:77–82.
23. Gill K, Blumenthal SL. Clinical experience with automated percutaneous diskectomy—the Nucleotome system. Orthopedics (in press).
24. Russell GS, Highland T, Dreisinger T. Percutaneous lumbar diskectomy: one-year follow-up in private practice. Surg Rounds Orthop (in press).
25. Swiecicki M. Results of percutaneous lumbar diskectomy compared to laminectomy and chemonucleolysis. In: Mayer HM, Brock M, eds. Percutaneous lumbar diskectomy. Berlin: Springer-Verlag, 1989:133–137.
26. Campanna AH, Campanna DM. Correlations of amount of disk removed by percutaneous lumbar diskectomy and clinical results. In: Mayer HM, Brock M, eds. Percutaneous lumbar diskectomy. Berlin: Springer-Verlag, 1989:138–140.
27. Cooney FD. Comparison of chemonucleolysis with chymopapain to percutaneous automated diskectomy: a surgeon's first 50 cases of each. In: Mayer HM, Brock M, eds. Percutaneous lumbar diskectomy. Berlin: Springer-Verlag, 1989:163–168.
28. Corkhill G, Kimple J, Corkhill R. Automated percutaneous nucleotomy by pain scale. In: Mayer HM, Brock M, eds. Percutaneous lumbar diskectomy. Berlin: Springer-Verlag, 1989:169–172.
29. Rezaian SM, Silver ML. Percutaneous diskectomy—personal observations of 27 cases. In: Mayer HM, Brock M, eds. Percutaneous lumbar diskectomy. Berlin: Springer-Verlag, 1989:173–176.
30. Cartolari R, Davidovits P, Gagliardelli M. Automated percutaneous lumbar diskectomy. In: Mayer HM, Brock M, eds. Percutaneous lumbar diskectomy. Berlin: Springer-Verlag, 1989:157–162.
31. Kaps H, Cotta H. Early results of automated percutaneous lumbar diskectomy. In: Mayer HM, Brock M, eds. Percutaneous lumbar diskectomy. Berlin: Springer-Verlag, 1989:153–156.
32. Davis GW, Helms C, Onik G. Automated percutaneous lumbar diskectomy: a prospective study of 518 cases with 1 year follow-up. Spine (in press).
33. Weigand H. APLD: report in 275 patients. German Yearbook of Radiology. Wiesbaden: 1991 (in press).
34. Kahanavitz N, Viola K, Goldstein T, Dawson E. A multicenter analysis of percutaneous diskectomy. Spine 1990; 15(7):713–715.
35. Jackson RP, Cain JE, Jacobs RR, Cooper BR, McManus GE. The neuroradiographic diagnosis of lumbar herniated nucleus pulposus: I. A comparison of computed tomography (CT), myelography, CT-myelography, discography, and CT-discography. Spine 1989; 14(12): 1356–1361.
36. Fries JW, Abodeely DA, Vijungco JG, Yeager VL, Gaffey WR. Computed tomography of herniated and extruded nucleus pulposus. J Comput Assist Tomogr 1982; 6:874–887.
37. Edwards W, Orme T, Orr-Edwards B. CT diskography: prognostic value in the selection of patients for chemonucleolysis. Spine 1987; 12:792–795.
38. Onik G, Maroon J, Shang Y. Far lateral disk herniation: treatment by automated percutaneous diskectomy. AJNR 1990; 11:865–868.
39. Onik G, Shang Y, Maroon JC. Automated percutaneous biopsy in postoperative diskitis: a new method. AJNR 1990; 11:391–393.
40. Onik G, Shang Y, Maroon J, et al. Percutaneous automated biopsy in the diagnosis of primary infectious spondylitis. Neurosurgery 1990; 26(2):234–237.
41. Onik G, Maroon J. Percutaneous automated diskectomy—a less invasive alternative for the treatment of herniated lumbar disks. Perspect Radiol 1988; 1(2):1–36.
42. Onik GM, Maroon J, Davis G. Automated percutaneous diskectomy at the L5–S1 level: use of a curved cannula. Clin Orthop 1989; 238:71–76.

# Arthroscopic Microdiskectomy

Parviz Kambin, M.D.

## Abstract

Arthroscopic microdiskectomy through a posterolateral approach has opened a new window of opportunity in the treatment of lumbar disk disorders. Radiographic identification of the triangular working zone has permitted the safe introduction of instruments with an external diameter of 7–8 mm into the intervertebral disk. The technique allows not only evacuation and decompression of contained herniated disks, but also the introduction of instruments for decortication of the vertebral plates and bone grafting for percutaneous interbody fusion. Endoscopic laser nucleolysis, currently under investigation, may also enhance existing technological achievement in the field of minimal-intervention spinal surgery. Arthroscopic microdiskectomy has proven to be safe, effective, and cost efficient. In properly selected patients, satisfactory results of approximately 85% have been realized.

ARTHROSCOPIC MICRODISKECTOMY through a posterolateral port represents an indirect approach to herniated lumbar disks. In contrast to laminectomy and microdiskectomy, the spinal canal is not entered and complications such as epidural bleeding, perineural fibrosis, and reherniation from the site of the anular fenestration are avoided.

The development of an indirect approach to the site of the herniation was motivated by the experience and published reports of complications and failures associated with the direct approach as described by Mixter and Barr (1). Microlumbar diskectomy (2) through a posteromedial approach does not avoid the undesirable complications associated with routine laminectomy procedure. In addition, striving for a limited incision and exposure may result in inadequate decompression or missing displaced sequestrated fragments.

The credit for the indirect approach to the lumbar intervertebral disk for the purpose of evacuation and decompression should go to Hult (3), who in 1956 described anterior retroperitoneal decompression of the disk for the treatment of sciatic pain. In 1963, Smith (4) also used the concept of indirect decompression of the intervertebral disk through chemonucleolysis. In 1973, I began my experimental work, which consisted of combining laminectomy for decompression of the spinal canal and introduction of a tube by the posterolateral approach to accomplish decompression and evacuation of a protruded disk (5).

Subsequently, clinical and cadaveric studies increased my confidence in employing negative pressure reaching as high as 735 mm of mercury. The description of the anatomic landmarks of the triangular working zone (6) and its radiographic appearance facilitated safe and proper positioning of the instruments. Better control and a certain measure of safety were provided by modified arthroscopes (7). Although laminectomy remains the cornerstone for the treatment of sequestered disks and spinal stenosis, arthroscopic microdiskectomy offers an alternative method of treatment for lumbar radiculopathy due to a herniated disk.

The advantages of the posterolateral approach include avoidance of epidural bleeding and perineural fibrosis, elimination of reherniation through the intraoperatively induced anular

---

From the Disc Treatment and Research Center, The Graduate Hospital, University of Pennsylvania School of Medicine, Philadelphia, PA. Address reprint requests to Parviz Kambin, M.D., Orthopaedic & Spine Surgery, Medical Institute, 1125 Lancaster Avenue, Berwyn, PA 19312.

fenestration, preservation of spinal stability, and establishment of a portal for future herniation away from the neural elements. Future surgical procedures are not compromised, and arthroscopic microdiskectomy is more cost effective because it decreases operating room time and length of hospitalization and allows the patient's early return to normal activities.

The role of conservative therapy for the management of radicular pain due to a protruded disk should not be ignored. The majority of patients respond to well-planned conservative therapy, including bed rest, nonsteroidal anti-inflammatory drugs or short-term steroid therapy, educational programs, physical therapy, and exercises. However, the treating physician should not be locked into prolonged and unreasonable conservative management.

Symptoms associated with intraneural and perineural fibrosis secondary to protracted mechanical pressure to the nerve root may be irreversible. If progression or development of a neurologic deficit occurs, surgical decompression should be considered. Finally, the patient's level of pain tolerance and demand for the use of narcotics should affect the timing of surgical intervention. The rehabilitation of a drug-dependent patient may be much more complicated than the slim chance of a surgical mishap.

## Indications

When surgical management becomes necessary, the physician has the responsibility to educate his or her patient about the available surgical modalities: lumbar laminectomy, posteromedial microlumbar diskectomy, arthroscopic microdiskectomy, and chemonucleolysis. In patient selection for arthroscopic diskectomies I have adhered to the following criteria (1):

- All patients initially had persistent sciatic pain and failed to respond to a course of conservative therapy.
- Many had one or two neurologic deficits.
- Some had correlative electromyographic evaluation.
- Mechanical signs, such as the straight-leg raising test, Lasègue's sign, or sitting root pain, were present.
- Myelographic, computed tomographic (CT), or magnetic resonance imaging (MRI) study confirmed the diagnosis.

In recent years, I have been able to depend on the combination of CT and MRI studies for localization of the symptom-producing disk. The need for routine myelographic studies has been eliminated.

Although, in certain centers, indirect decompression of the intervertebral disk is being utilized for the treatment of lumbago, diskogenic pain, and internal disk destruction syndrome, I do not subscribe to such management.

Patients with sequestered fragments, previously operated-on disks, failed chemonucleolysis, or spinal stenosis, elderly patients with chronic degenerative bulging disks, and individuals with cauda equina syndrome are not candidates for arthroscopic microdiskectomy. As far as the type of herniation is concerned, I have treated both central and posterolateral herniations successfully. The size of the herniation also should not be a major concern. I have decompressed and partially evacuated rather large contained subligamentous fragments; proper positioning of the sheath is particularly important for the evacuation of such herniations.

## Operative Technique

The surgical instruments for this procedure have undergone further refinement. However, the original universal working sheath with an outer diameter of 6.5 mm, providing 5-mm working access, is still in use. The modified versions of arthroscopic tools—diskoshavers and suction punch forceps—permit adequate and rapid nuclear extraction.

To reduce postoperative morbidity, I have performed most diskectomies through a uniportal approach. However, a biportal approach has sometimes been employed.

To diminish the chance of infection and its serious consequences, I strongly feel that the procedure should be performed in the operating room with a strict sterile environment. I administer prophylactic antibiotic for 24 hr to all my patients.

**Positioning the Patient and the Fluoroscope.** The procedure is performed with the patient in the prone position, which minimizes any movement or the need to reposition the trunk during surgery and prevents false X-ray imaging in the anteroposterior and lateral projections. A radiolucent, adjustable frame allows freedom of the abdomen and flattens the lumbosacral angle. Proper positioning is particularly important for L5–S1 disk extraction.

The C-arm fluoroscope is positioned and covered by a sterile plastic sheet. Blockage of rotation of the C-arm should be avoided to assure a reproducible anteroposterior and lateral image. For better visualization of the L5–S1 disk space, a slight tilt of the C-arm may be necessary. A marker is placed on the skin to confirm the surgical level in the anteroposterior projection.

**Anesthesia.** The procedure must be performed with the patient under local anesthesia; intraoperative communication with the patient is essential. Local anesthesia is obtained with 1% lidocaine solution; the skin and subcutaneous tissue are infiltrated. Infiltration of the paraspinal muscles with lidocaine containing epinephrine solution minimizes the incidence of postoperative psoas hematoma. Avoiding perianular infiltration prevents anesthetizing the nerve root and thus predisposing it to injury. The point of entry is 9–10 cm from the midline on the symptomatic side. However, a distance of 8–9 cm may be adequate for slim patients.

**Positioning the Needle.** Correct positioning of the needle is the single most important step affecting the final outcome of the operative procedure. The angle of introduction of the needle may be predetermined by an abdominal CT at the level of the herniated intervertebral disk. Generally, the angle of introduction of the needle is about 35° to 45°. The inserted needle should penetrate the sacrospinalis, quadratus lumborum, and psoas major muscles in its path. When the needle is introduced lateral to the boundary of these muscles, there is a greater likelihood of entrance into the abdominal cavity and of serious complications. In contrast, entry too close to the midline may impede adequate evacuation.

The tip of the needle should approach the anulus just lateral to the proximal articular process of the inferior vertebra within the triangular working zone. In this position, the tip of the needle is visualized in the lateral projection on the posterior border of the intervertebral disk, or slightly anterior. In the anteroposterior projection, the tip of the needle is seen in alignment or just lateral to the midportion of the pedicles. Radiographic documentation is done while the tip of the needle is resting on the anular fibers. Premature introduction of the needle into the intervertebral disk will alter the radiographic appearance and should be avoided.

As the needle is inserted, the surgeon must rely on hand-eye coordination. Resistance should be encountered when the needle reaches the anulus.

If no resistance is encountered and the lateral projection demonstrates that the needle has bypassed the anulus, a verticle insertion has been performed. An anteroposterior X-ray would confirm that the needle has not reached the anulus. Withdrawal of the needle and reinsertion at a more horizontal angle is indicated.

In contrast, if the needle has been inserted at an excessively steep horizontal angle, resistance is usually encountered but the lateral X-ray would demonstrate that the tip of the needle has not reached the anulus. Withdrawal of the needle and reinsertion at a more vertical angle is indicated.

In the course of anatomic dissection, when the dorsolateral approach is used and the instruments are inserted parallel to the vertebral plates, the spinal nerve is protected for the most part from the instruments by the facet joints, pedicles, and transverse processes. Only a small portion of the spinal nerve situated between the foramina and the superior border of the transverse process is subject to penetration and insult. At the site of anular fenestration, the spinal nerve is separated from the anulus by a thin layer of fatty tissue and the fibers of the psoas major muscle.

As the spinal nerve descends across the intervertebral disk, it creates a triangular working zone, which is bordered by the spinal nerve, the proximal plate of the inferior vertebrae, and the proximal articular process of the lower lumbar segment.

**Placing Guide Wire, Trocar, and Sheath.** Following the proper positioning of the needle, the stylet is replaced by a 0.028-in. Kirschner wire. The needle is then withdrawn. As the guide wire enters the anulus, the surgeon should experience resistance. At this time, a specially designed, cannulated, blunt trocar is passed over the Kirschner wire (Fig. 1). The surgeon should watch for possible migration of the guide wire in the course of introduction of the cannulated trocar.

As soon as the direction of the cannulated trocar is established, the guide wire should be removed. This precautionary measure ensures that the blunt end of the trocar will bypass the nerve root and prevents undue damage to the nerve fibers. The universal sheath with an external diameter of 6.5 mm is passed over the trocar until it reaches the anulus fibrosus (Fig. 2).

The suction irrigation adaptor is now attached to the working sheath. The diskoscope is introduced, the suction irrigation system with normal saline solution is established, and the anulus is inspected (Fig. 3). The outflow tube is

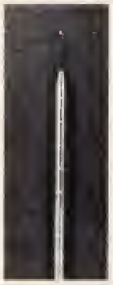


**Fig. 1.** Cannulated trocar allows precise insertion of instruments and follows path of injected local anesthetics.

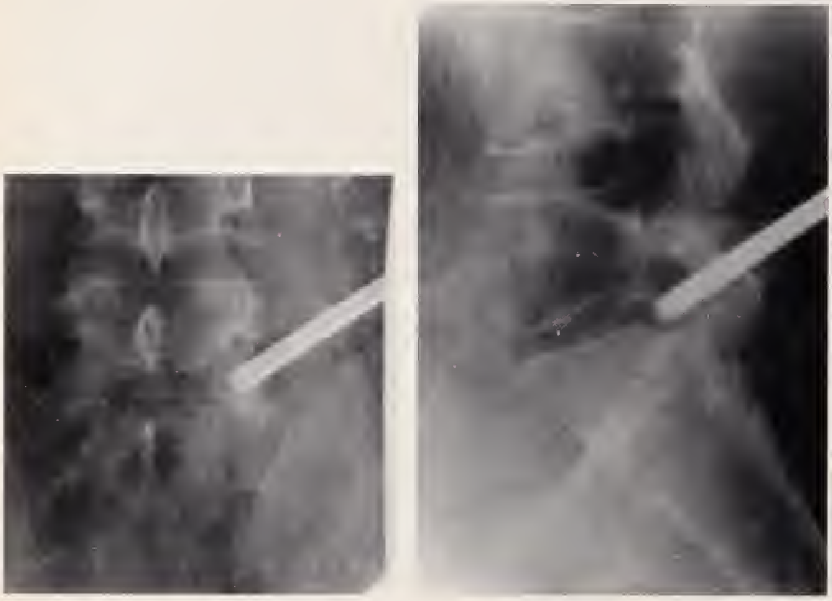


**Fig. 2.** Intraoperative fluoroscopy. Anteroposterior projection. Note proper position of working sheath for decompression of L5–S1 disk with distal extremity of sheath in alignment with midpart of pedicle. Lateral projection. Note sheath in position just posterior to anulus.

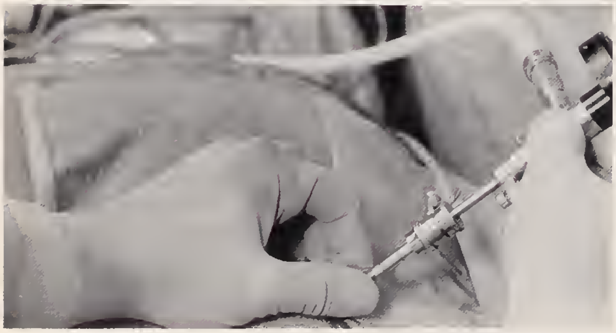
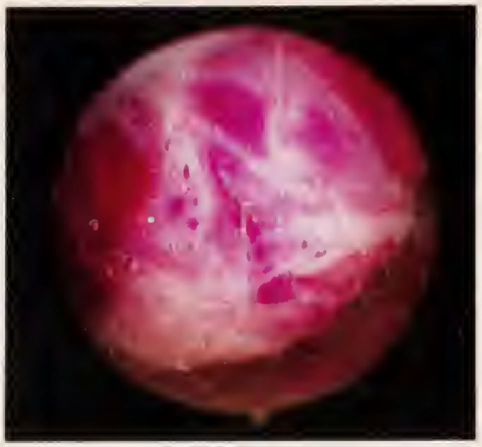


**Fig. 3.** Suction irrigation adaptor is attached to working sheath for arthroscopic inspection of anulus and perianular structures.

connected to the suction irrigation adaptor, and the inflow tube is attached to the arthroscope irrigation sheath for better visualization. The suction is turned down to 125 mm–255 mm of mercury.

**Confirming Absence of Neural Tissue.** When the working sheath is properly positioned in the triangular working zone, the surgeon will invariably observe loose adipose tissue covering the anulus. At times, a thin layer of connective tissue is seen over the anular fibers. For better inspection of the anulus, the adipose tissue may be removed with forceps or wiped off with a cottonoid. Inspection of the anulus will unequivocally confirm the absence of neural tissue under the open end of the working sheath (Fig. 4).

A simple needle test may also confirm the absence of nerve root entrapment. Needling of the anulus by walking around the internal diameter of the sheath should not produce radicular pain. However, if the nerve is trapped, introduction of the needle will cause severe radicular



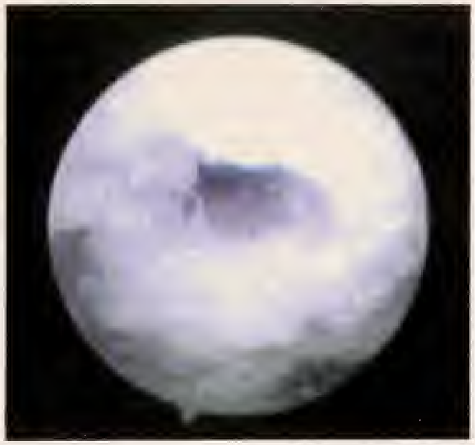


**Fig. 4. Left** Arthroscopic view of anulus following removal of adipose tissue. **Fig. 6. Right** Cottonball appearance of nucleus. Note cavity created behind anulus at site of herniation.

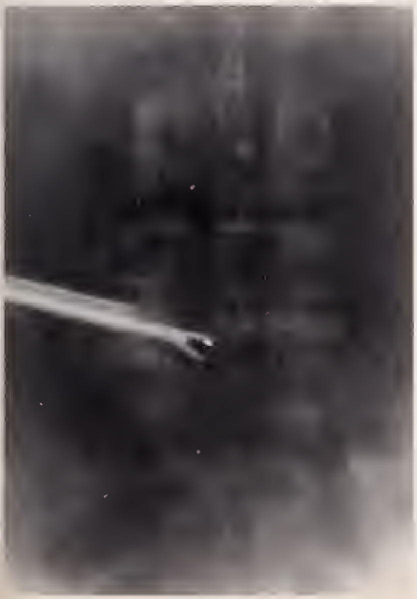
symptoms. Needle testing is also used prior to fenestration of the anulus to ensure that the open end of the working sheath is not resting on the vertebral plate of the adjacent segments. If the needle meets bony resistance, the sheath is tilted until it is well centered on the anulus fibrosus.

**Fenestration of Anulus.** The suction irrigation adaptor is removed. Because fenestration of the anulus is usually painful, additional systemic analgesics may be necessary. Infiltration of the anulus fibers with local anesthetics is difficult, particularly in younger individuals. However, topical anesthesia may be obtained by placing a cottonoid saturated with lidocaine solution on the anulus.

Fenestration is accomplished first with a small and then with a larger circular cutting instrument by a firm rotatory movement maintaining downward pressure.

**Limiting Penetration to 2 cm.** Following fenestration of the anulus, the posterior nucleus is evacuated with a forceps (Fig. 5), suction, and power tools. The suction irrigation adaptor must be reattached to the working sheath prior to the introduction of forceps or powered diskotomes. This added measure of safety limits the depth of penetration of the instruments to 2 cm.

**Evacuation.** Evacuation of nuclear material from the center of the intervertebral disk is not enough. The challenge for successful arthroscopic microdiskectomy lies in the ability to reach posteriorly and to evacuate the interanular or subligamentous offending fragments. A curved, back-biting forceps is employed to reach and



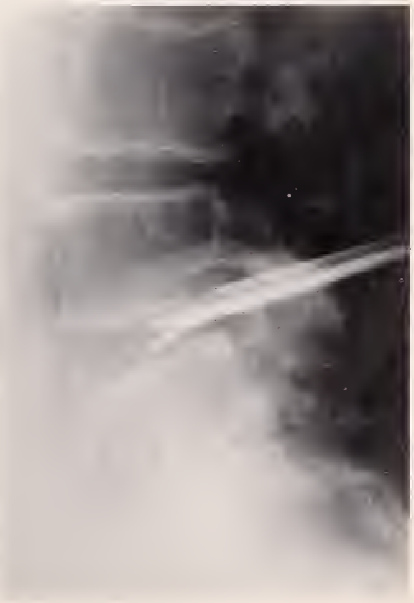


**Fig. 5.** Anteroposterior (**left**) and lateral (**right**) fluoroscopy showing micropituitary forceps at L4–L5 intervertebral disk.

evacuate these fragments from an area just adjacent to the tip of the inserted sheath. The sheath is then engaged and advanced into the anulus for a distance of 2 or 3 mm. Then the procedure is repeated. The surgeon should guard against advancing the sheath too far into the intervertebral disk. The position of the tip of the forceps should be checked periodically with the fluoroscope.

A 50-mL syringe or suction pump, commonly used for liposuction, is capable of generating about 740 mm of mercury negative pressure. Creation of a vacuum in the center of the intervertebral disk brings the loose nuclear fragments into the path of the sheath and the inserted instruments.

The deflector tube and flexible forceps are also useful for extraction of the posterior fragments.

A final inspection of the operative site will ensure that an adequate cavity has been created behind the posterolateral anulus and that no free fragment is left behind (Fig. 6).

## Conclusions

Arthroscopic microdiskectomy has proven to be nondestructive, safe, effective, and cost-efficient. In properly selected patients, a satisfactory result in 87% may be expected (5). The earlier concern of intraoperative complications, specifically the development of neural injury, seems to be unfounded. The success of the procedure hinges on the following principles:

- Knowledge of the anatomy of the spinal nerve and its surrounding structures
- Adherence to every detail of the operative technique
- Careful isolation of the symptom-producing disk
- Adequate anular fenestration
- Correct positioning of the working sheath in the triangular working zone, just lateral to the proximal articular process of the lower lumbar segment
- Evacuation of herniated disk material from an area adjacent to the open end of the working sheath as it rests on the outer surface of the anulus outside the foramina

Certainly, proper patient selection and recognition of the effect of psychosocial elements in the persistence of pain remains the single most important factor affecting the final outcome of arthroscopic microdiskectomy.

### References

1. Mixter WJ, Barr JS. Rupture of the intervertebral disc with involvement of the spinal canal. N Engl J Med 1934; 211:210.
2. Love JG. Removal of protruded intervertebral discs without laminectomy. Proc Mayo Clin 1939; 14:800.
3. Hult L. Retroperitoneal disc fenestration in low back pain and sciatica. Acta Orthop Scand 1956; 20:342.
4. Smith L. Enzyme dissolution of the intervertebral disc. Nature 1963; 4887:198.
5. Kambin P, Gellman H. Percutaneous lateral discectomy of the spine. Clin Orthop 1983; 174:127–132.
6. Kambin P, Schaffer JL. Percutaneous lumbar discectomy: prospective review of 100 patients. Clin Orthop Rel Res 1989; 1:24–34.
7. Kambin P, ed. Arthroscopic microdiscectomy, minimal intervention in spinal surgery. Baltimore, MD: Urban & Schwarzenberg, 1991.

# Minilaminotomy as an Alternative to Laminectomy or Microdiskectomy: Ten Years' Experience

Martin H. Savitz, M.D.

## Abstract

Minilaminotomy as an alternative to microdiskectomy or standard laminectomy has been performed at two community hospitals on 1000 patients over a period of 10 years. The operative technique, employing simple prone positioning, a 4-cm incision, standard instruments, X-ray guidance, a fiberoptic head light, and 4.5× binocular loupes, has not changed since 1986, when the initial series of 200 cases was published. The overall results in 800 additional patients continued to compare favorably to all reported series of microdiskectomy: minimal blood loss, less narcotic requirement, fewer days in the hospital, earlier return to normal activity, no wound infection, and low rate of second surgery.

In 1932, Barr (1) was the first to understand that chondromata, cartilaginous tumors excised during decompressive laminectomy, were actually material herniated from an intervertebral disk space. As early as 1939, Love (2) proposed small laminotomy incisions with limited removal of the protruding or extruded nucleus pulposus. The first neurosurgeon to regularly use the binocular microscope in the operating room was Malis (3), who began to perform microsurgical diskectomy in 1968, when only Pantopaque myelography was available. In 1972, Williams (4) had unique motivation to employ microlumbar diskectomy, namely Las Vegas showgirls who could not display more than a one-inch incision or stay out of work more than a few weeks. The term "minilaminotomy" was coined to describe microsurgical techniques which not only relieved segmental spondylosis and lateral recess syndrome but also provided adequate exposure of sequestered fragments (5).

## Methods and Materials

The operation, employing simple prone positioning, a 4-cm incision, standard instruments, X-ray guidance, a fiberoptic head light, and 4.5× binocular loupes, has already been described (Fig.), but certain techniques were not emphasized in the initial report (5). During the opening, fat was not only obtained for grafting, but the subcutaneous layer was cut out when necessary to avoid interference with the light source. Proper illumination of the surgical field usually required resection by Leksell rongeur of all muscle not completely retracted by the Taylor retractor. A second X-ray on the table confirmed the level of exploration and revealed that the disk space was superior to the interlaminar space in most cases. Opening the anulus fibrosus on one side even after medial facetectomy gave an exaggerated sense of laterality when either herniation of the nucleus pulposus into the foramen or a laterally retained fragment within the disk space was visualized on computed tomography (CT) or magnetic resonance imaging (MRI). Although curettage of the cartilaginous plates was specifically avoided, removal of 6–10 gm of disk material was possible with pituitary rongeurs

---

From the Division of Neurosurgery, Department of Surgery, Good Samaritan Hospital and Nyack Hospital, Rockland County, New York. Address reprint requests to Martin H. Savitz, M.D., 55 Old Turnpike Road #101, Nanuet, NY 10954.

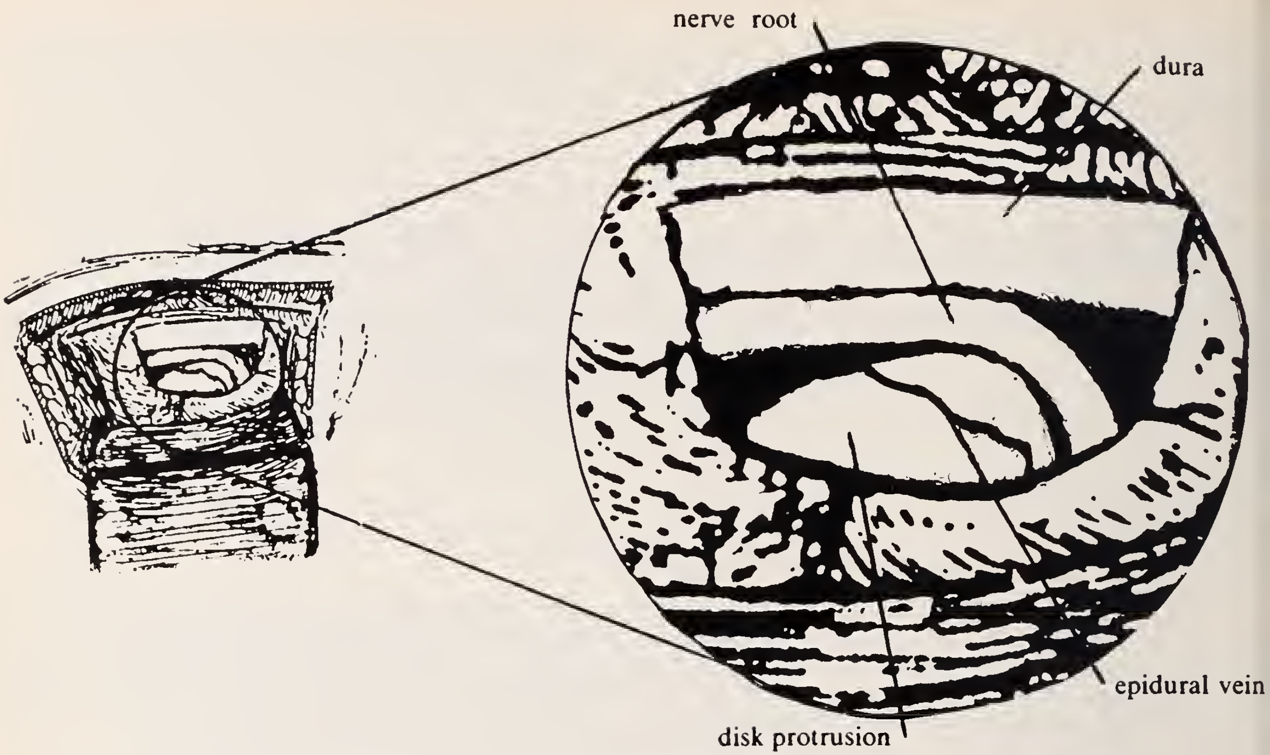


**Fig.** Artist's drawing of minilaminotomy incision with surgeon's view magnified 4.5× by binocular loupes.

and by direct excision of scarred-in fragments with a no. 11 scalpel. Erosion of the dural sleeve of the nerve root with leakage of cerebrospinal fluid was observed in 17 patients; direct repair with 5-0 Nurilon in 12 patients and patching with Gelfoam in the remaining 5 prevented postoperative fistulae.

## Results

The ages of the patients ranged from 14 to 82. A total of 512 men and 488 women came to surgery; there were 31% compensation cases, and 134 patients had undergone prior diskectomy. The L5–S1 disk was herniated in 492 patients, the L4–5 disk in 388 patients, the L3–4 disk in 30 patients, the L2–3 disk in 5 patients, both L3–4 and L4–5 in 20 patients, and both L4–5 and L5–S1 in 65 patients. No complications of increased neurologic deficit, wound infection, or hemorrhage requiring transfusion were recorded in the series of 1000 minilaminotomies.

Two instances of thrombophlebitis were the only complications reported in the initial 200 operations. Three other complications occurred in the subsequent 800 procedures. A 38-year-old man fractured the L4 facet when he got out of bed on his second postoperative day; immediate reexploration was performed to remove the bone fragment and alleviate his recurrent sciatica. One instance of diskitis occurred in a 45-year-old woman who was reexplored for scarring and segmental spondylosis at the L4–5 level; no organism was cultured from three separate aspirates of the disk space. One postoperative cerebrospinal fluid fistula developed at the L4–5 level in a 52-year-old man who had undergone diskectomy at L4–5 and L5–S1; no leak was observed during his initial surgery, but reexploration was required for direct repair.

Elaborate methods of assessing the results of lumbar disk surgery have been described (6). Questionnaires, psychological testing, serial neurological examination, the bias of compensation cases, electromyographic testing, and computer analysis have all been applied. A more direct way was simply to ask each patient to judge the success of the operation. To the question, "Are you better?" 922 replied affirmatively, and 78 said, "No."

The patients who were discharged home but were not sufficiently improved to return to work were reevaluated by MRI with gadolinium rather than CT, as previously reported (5); the neuroradiologic imaging revealed postoperative changes at the correct level. Seventeen patients were reexplored; the fat graft made the dissection technically simple. The operative findings in 15 of these patients were scar tissue arising from the area of the incised anulus and attaching to the nerve root. Lysis of the adhesions was performed, but only 7 patients reported diminished radicular pain. Two patients who were reoperated on had significant granulation tissue and some disk fragments superior to the level of initial exploration. Two patients underwent radiofrequency facet rhizotomy (7), and both improved. Twenty of the 800 patients who failed to gain any relief of pain were diabetic.

## Discussion

Williams (4) advocated (a) preoperative X-ray localization with as small a skin incision as possible, (b) the magnification and illumination provided by the operating microscope, (c) no alteration of the lamina or facet joint, (d) no coagulation or removal of fat within the epidural space, and (e) removal of a minimal volume of disk material without incision of the anulus or curettage. Subsequent clinical investigation (8–11) concluded that any one of these principles might be responsible for results superior to standard laminectomy.

Positioning of the patient either laterally or obliquely permits the surgeon to sit while using the microscope, but does not allow intraoperative confirmation of the disk exploration; using only preoperative X-rays, the correct level may be missed (3). Paramedian incision was stressed by Wilson and Harbaugh (8) for placement of the Williams retractor, and by Balagura (10) to avoid cutting the supraspinal ligament. The incision for minilaminotomy was midline and 4 cm in length rather than 1 inch. An ordinary Taylor retractor hooking on the facet provided complete retraction of the paraspinal muscle and fascia.

No attempt was made to avoid removal of bone or ligamentum flavum, as suggested by Williams (4); standard Kerrison punches were not abandoned in favor of a high-speed drill. Laminotomy was necessary in most patients to visualize the superiorly positioned disk space. Medial facetectomy and foraminotomy not only relieved segmental spondylosis and lateral recess syndrome but facilitated removal of extruded fragments without excessive nerve root retraction. As for the epidural fat and veins, hemostasis and clear inspection of the posterior longitudinal ligament were the prime considerations. Not scraping the cartilaginous plates with curettes may lessen the chances of diskitis.

Minilaminotomy obviates extensive dissection and exploration of surrounding anatomy while facilitating visualization of the immediate environs of the compressed nerve root. Preoperative neuroradiologic studies and intraoperative X-rays allow exact localization and precise removal of a herniated nucleus pulposus through a modest incision. The size of the surgical wound, dimensions of the bone opening, quantity of disk material excised, and amount of perineural fat preserved are not as important as the skill of the surgeon, correlation of neuroradiologic images with clinical findings, patient selection, and avoidance of complications.

## References

1. Barr JS. Lumbar disc lesions in retrospect and prospect. Presented at San Diego Naval Hospital, San Diego, California, May 15, 1961.
2. Love JG. Removal of the protruded intervertebral discs without laminectomy. Proc Mayo Clin 1939; 14:800.
3. Sachdev VP. Microsurgical lumbar discectomy: a personal series of 300 patients with at least 1 year of follow-up. Microsurgery 1986; 7:55–62.
4. Williams RW. Microlumbar discectomy: a conservative surgical approach to the virgin herniated lumbar disc. Spine 1978; 3:175–182.
5. Savitz MH, Katz SS. Minilaminotomy, an alternative to microdiskectomy or standard laminectomy, Mt Sinai J Med 1986; 53:518–522.
6. Rish BL. A critique of the surgical management of lumbar disc disease in a private neurosurgical practice. Spine 1984; 9:500–504.
7. Katz SS, Savitz MH. Percutaneous radiofrequency rhizotomy of the lumbar facets. Mt Sinai J Med 1986; 53:523–525.
8. Wilson DH, Harbaugh R. Microsurgical and standard removal of the protruded lumbar disc: a comparative study. Neurosurgery 1981; 8:422–427.
9. Spengler DM. Lumbar discectomy: results with limited disc excision and selective foraminotomy. Spine 1982; 6:604–607.
10. Balagura S. Lumbar discectomy through a small incision. Neurosurgery 1982; 11:784–785.
11. de Divitiis E, Spaziante R, Cappabianca P, Donzelli R. Lumbar disk: surgical tricks for safeguarding the "root's ecology." Surg Neurol 1984; 22:73–75.

# Spinal Instrumentation for Degenerative Disease of the Lumbar Spine

MICHAEL J. HARRISON, M.D., AND NARAYAN SUNDARESAN, M.D.

## Abstract

Instrumentation for spinal implant is currently changing more rapidly than at any previous period in its 100-year history. At present, the most common indication for spinal instrumentation is degenerative disease of the lumbar spine, which can be roughly categorized as of three kinds: herniated lumbar disks, spondylolisthesis, and spinal stenosis. The literature on indications for arthrodesis is reviewed. A brief summary of the approach and type of instrumentation available for the lumbar spine is presented.

ALMOST 100 years have elapsed since the initial attempts at spinal instrumentation. The use of wire and rods in spinal stabilization was first reported in the English literature in the late 1800s and early 1900s (1, 2). These articles slightly predate the documentation of spinal fusion by Albee (3) and Hibbs (4) in 1911. The combination of fusion and internal fixation was not developed until King (5) used pedicle screws to achieve arthrodesis during lumbosacral fixation in 1944. During the 1960s, Harrington (6) developed instrumentation for correction of paralytic scoliosis following poliomyelitis. This system has become the standard by which all others are measured. Surgical arthrodesis has undergone profound changes since these initial papers, yet never before the past decade has such rapid evolution of spinal instrumentation occurred (7–17).

Throughout the history of arthrodesis and instrumentation, pseudoarthrosis has remained a major complication (7). If bony fusion does not occur, instrumentation is doomed to failure. Without bony fusion, dislodgment or fracture of the implant will occur. The success of the implant is contingent on the success of the arthrodesis. The three objectives of spinal instrumentation are (1) to provide an ideal rigid milieu in which bony fusion can take place, (2) to provide load sharing with the spine to prevent fracture and breakdown while arthrodesis is occurring, and (3) to restore alignment to the spinal axis.

### Stability and Instability

If stability is the goal of spinal fusion, then one must be able to define instability. Instability, as defined by White and Panjabi (8), is "the loss of the ability of the spine under physiologic loads to maintain its pattern of displacement so that there is no initial or additional neurologic deficit, no major deformity, and no incapacitating pain." Although eloquently worded and addressing the fundamental issues of pain, neurologic deficit, and deformity, this statement is not a clinically practical definition applicable to patient selection for surgical intervention.

Numerous articles have addressed the issue of instability in the lumbar spine (8, 18–25). Contrary to most experts, Paris (26) asserts that instability can be seen, palpated, and distinguished from hypermobility on physical examination. X-rays with flexion and extension views have almost universally been employed to define instability. Translational movement in an anteroposterior plane of 4–5 mm (8%–9%) is generally acknowledged as excessive (27–31). The difficulty with defining instability this way is that

---

From the Department of Neurosurgery, The Mount Sinai Hospital and Mount Sinai School of Medicine (CUNY), New York, NY. Address reprint requests to Narayan Sundaresan, M.D., Department of Neurosurgery, Box 1136, The Mount Sinai Hospital, One Gustave L. Levy Place, New York, NY 10029.

TABLE

*Checklist for Clinical Instability in the Lumbar Spine*

| Clinical characteristic | Points* |
|---|---|
| **Elements destroyed or nonfunctional** | |
| Anterior | 2 |
| Posterior | 2 |
| **Radiologic criteria** | |
| Flexion-extension X-rays | |
| Sagittal translation >4.5 mm or 15% | |
| Sagittal plane rotation | |
| >15% at L1–2, L2–3, L3–4 | 2 |
| >20% at L4–5 | |
| >25% at L5–S1 | |
| or | |
| Resting X-rays | |
| Sagittal plane displacement >4.5 mm or 15% | |
| Relative sagittal plane angulation >22% | |
| **Cauda equina damage** | 3 |
| **Dangerous loading anticipated** | 1 |

* Total of 5 or more points = unstable.
*Adapted* from White and Panjabi (ref. 8).

asymptomatic patients are frequently encountered with the same degree of movement. Further difficulties lie in the fact that instability can be the result of lateral translation and rotational instability. Other radiographic findings such as spondylolisthesis, disk space narrowing, disk space gas (Knutsson's sign), facet hypertrophy, and traction spurs may indicate instability (24, 31, 32). Computed tomography scans obtained with the patient supine and sandbags placed beneath one hip may also be helpful in showing opening of the contralateral facet joint and encroachment on the ipsilateral neural foramen with elicitation of symptoms (24).

Some clinicians have touted the merits of facet blocks and diskography in the diagnosis of instability (19, 33); results have been mixed, and predictive capabilities are open to dispute (34).

Although external orthosis has long been used to ascertain lumbar spinal instability, a promising diagnostic development is external transpedicular fixation (34, 35), which utilizes pedicle screws inserted percutaneously under general anesthesia with fluoroscopic guidance. A frame is then attached externally to the screws to rigidly fix certain segments of the spine. In the most recent series (34), 35 patients were prospectively studied; results showed a statistically significant superiority of temporary external spinal fixation in predicting the surgical outcome of fusion as compared with plain radiographs, diskography, and facet blocks. Trials of external fixation have only been performed within the socialized systems of medicine in Sweden and Canada.

To date, no universally acceptable criteria for patient selection are available that ensure success of lumbar arthrodesis in the treatment of low back pain disorders (7, 36, 37). Although checklists such as that provided by White and Panjabi (8) may be useful (Table), each physician must rely on personal experience and the surgical literature.

## Lumbar Disk Disease

Although the actual surgical technique for the removal of lumbar intervertebral disks has changed very little since its introduction by Mixter and Barr (38) in 1934, there has been much debate over the need for accompanying fusion. Fusion in association with disk excision is indicated only when the spinal segment is already unstable (39, 40). Diskectomy by fenestration and minimal resection of the lamina does not produce instability (8). Concern has been expressed that removal of a lumbar apophyseal joint, particularly in association with removal of the lumbar intervertebral disk, may produce instability. Hazlett and Kinnard (41) evaluated 33 patients clinically and radiologically 2–5 years after lumbar nerve root decompression by apophyseal joint excision and diskectomy. A one-level unilateral or bilateral resection of the facet joint in association with diskectomy does not result in segmental instability. The four patients in the study who developed instability had two or three apophyseal joints resected.

However, Panjabi (42) has contrary feelings on the issue. Based on experiments performed in cadaver spines, he concluded that partial facetectomy of one or both facets at a single level does not cause spinal instability, whereas the loss of a complete facet joint on one or both sides makes the spine acutely unstable.

In 1952, Nachlas (43) reported a 5-year follow-up on patients under 45 years of age who underwent 256 simple disk excisions and 118 disk removals with fusion. Within the fusion group there was a longer hospital stay, a longer period before resuming work, and a greater complication rate. The study concluded that when surgical intervention is necessary for simple cases of herniated nucleus pulposus, the operation of choice is disk excision alone. Frymoyer et al. (44) evaluated 244 patients 10 years following either disk excision or diskectomy with fusion (H-graft or Hibbs); statistical analysis concluded that midline spinal fusion offered no benefit in the management of lumbar disk disease.

White et al. (45) in 1987 published a controlled prospective study on a matched set of 69 patients with herniated lumbar disks. Both groups underwent bilateral laminectomy and disk excision by the same surgeon. One group had the addition of an intertransverse fusion with internal fixation. Among the fusion group, the mean time to resuming work was significantly longer. The success rate was approximately equal (87%), but the best results were obtained in the nonfusion group. Twenty-nine percent of the nonfusion group had excellent results, versus 11% of the fusion group. The conclusion was that fusions were unnecessary, produced less excellent results, and led to a longer recuperative period. Spinal fusion plays no role in the management of simple herniated nucleus pulposus and should be restricted to the management of spinal instability.

## Spondylolisthesis

Killian (32) is credited with the first description of spondylolisthesis in 1853. Spondylolisthesis is now recognized to arise from multiple etiologies (32, 46, 47):

- congenital (dysplastic)
- isthmic (spondylolytic)
- degenerative
- post-traumatic
- pathologic
- post-surgical

Our discussion is limited to spondylolytic and degenerative spondylolisthesis.

**Spondylolytic Spondylolisthesis.** Spondylolytic or isthmic spondylolisthesis is by far the most common type (48). Whether the defect in the pars interarticularis results from a congenital defect or a stress fracture is beyond the scope of this article. Genetic factors undoubtedly play a role in some cases, but the majority of the data supports stress fracture as the underlying etiology (32). Prior to the report of Gill et al. (49) documenting successful treatment by decompression alone, the mainstay of therapy for spondylolytic spondylolisthesis was fusion. In 1965, Gill and White (50) presented a 10-year follow-up on 43 patients treated by decompression with an 86% success rate. Two salient observations were made: first, only patients 25 years of age or less experienced progression of slippage postoperatively, and, second, there was no correlation between postoperative progression and outcome.

In 1970, Amuso et al. (51) reported on 33 patients with a mean follow-up of 7 years following the Gill procedure for a grade I or II spondylolytic spondylolisthesis; there were only 64% satisfactory results. Among the 36% failures, half the patients presented with low back pain alone and no signs of nerve root compression. Only three patients had postoperative slippage of more than 10%, but again, progression bore no relation to final results. What can be gleaned from this series is that the Gill procedure is inappropriate for patients with back pain and that the people who benefit from posterior element resection demonstrate preoperative evidence of nerve root compression.

More recently, Österman et al. (52) documented 75 patients with a minimum of 5 years follow-up after Gill's operation; there were 75% long-term satisfactory results. Most of the excellent results were in younger patients, whereas most of the unsatisfactory results were encountered in older patients who had degenerative disk disease. The degree of primary displacement had no effect on the final results, and postoperative progression of slippage had no bearing on outcome.

What can be surmised from the literature is that Gill's operation is the procedure of choice for patients with signs of nerve root impingement. In patients with back pain alone, decompression will not help and fusion is indicated. Although progression of slippage does not correlate with outcome, a fusion should be considered in younger patients undergoing posterior element resection. Progression has been documented primarily in patients under 25 years of age, but a more conservative approach would be to include fusion in patients 40 years and younger. When disk disease is present, without surgical correction any other procedure alone will fail.

**Degenerative Spondylolisthesis.** Degenerative spondylolisthesis results from long-standing segmental instability. It is found 4 to 6 times more often in females, 3 times more commonly in blacks, and 6 to 10 times more frequently at the L4 level than at L3 or L5 (47, 53). Because the disease is found primarily in elderly patients, osteoarthritic changes of the articular apophysis are usually severe (32, 48, 53). The listhetic deformity is not as severe as in isthmic spondylolisthesis and slipping generally does not exceed 30% (32, 47, 53).

By combining the series of Newman (48) and Rosenberg (53), over 280 patients with degenerative spondylolisthesis may be analyzed. The vast majority of patients may be managed conservatively, and only about 15% will come to surgery. Indications for surgical intervention are: (1) persistent severe back pain unresponsive to conservative treatment, (2) persistent sciatica, and (3) the syndrome of cauda equina claudication compression. Pain in the low back, buttock, or thigh was the principal complaint bringing patients to medical attention. the pain was usually longstanding (up to 5 years), and a few patients had only thigh pain. Intermittent claudication was uncommon but usually presaged the need for surgical decompression. More than 25% of the patients had symptoms in the leg or foot, usually sensory in nature and most often involving the fifth lumbar nerve root. Cauda equina symptoms such as diminished sphincter tone or footdrop were present in less than 4% of the patients.

In Rosenberg's (53) series of 30 surgically treated patients, five had X-ray evidence of instability postoperatively, yet radiological slippage was not consistently present in patients troubled by symptoms of instability. The progression of slippage was as much as 15% in one patient and was complete by the end of one year after surgery in all cases.

Adequate decompression is of paramount importance in surgical management. Partial bilateral facetectomy is sufficient in most cases to decompress the nerve roots and leave sufficient bone laterally to prevent future progression of the listhesis. But as decompression is of the essence, it is not always possible to leave the facets laterally. Removal of the neural arch may also be necessary, posing the potential problem of instability. Once decompression is complete, a decision can be made on whether to surgically stabilize. In 1985, Feffer et al. (54) compared two matched populations, one undergoing decompression accompanied by fusion and the other decompression alone for degenerative spondylolisthesis. Although the series were small, all 8 patients receiving fusion had good or fair results, but 3 of the 11 patients with decompression alone had poor results. Newman (48) and Rosenberg (53) believed that fusion is rarely necessary and should be left to a secondary procedure, but Feffer et al. (54) recommended routine fusion in all cases. Herron and Trippi (55) studied 24 patients with an average follow-up of 34 months following laminectomy and partial facetectomy and without fusion; the success rate was close to 96%. In the absence of objective instability on preoperative flexion-extension X-rays, fusion was not required when the decompression preserved the integrity of the pars interarticularis and the articular processes.

No verdict has been reached on whether to incorporate fusions routinely into all surgery for degenerative spondylolisthesis. Because radiologic progression of the listhesis does not correlate with symptomatic instability postoperatively, it is difficult to ascertain what objective findings the physician can follow. For a patient with back pain alone, fusion is the treatment of choice. When signs of nerve root compression are present, thorough decompression is mandatory. If the lateral half of the facet joints can be spared, fusion may be reserved for a secondary procedure. When radical removal of the facets is required, fusion should proceed, particularly in patients under 65 years of age. Fusion is by no means a benign procedure. If fusion were routinely recommended for all patients, the operative time would be increased in elderly patients, the recuperative period would be prolonged, and the rate of complications would be increased. In addition, long-term pain at the donor site is common.

## Spinal Stenosis

Spinal stenosis and degenerative spondylolisthesis may be considered part of the same disease process—chronic segmental instability. In fact, degenerative spondylolisthesis is frequent in patients who have multilevel spinal stenosis. Both entities occur in older patients; the average age is above 60 years in most series. As in spondylolisthesis, adequate decompression is the primary surgical objective.

Shenkin and Hash (56) in 1979 reported a series of 59 patients with 6 years' follow-up of multiple-level laminectomy and facetectomy for lumbar spondylosis. At least a two-level operation including L4 and L5 was performed in all patients, and an operation up to four levels in some patients; overall, 89% did well. Postoperative X-ray studies of patients who fared poorly failed to reveal a correlation with slippage. Six of 59 patients (10%) slipped, yet 4 of them did well despite the slippage. The average age of patients with slippage was somewhat younger than the average of the series as a whole. In cases of two-level laminectomy and facetectomy, there was a 6% incidence of postoperative slippage. However, when three or four levels were decompressed, the rate of slippage was 15%. When laminectomy and facetectomy is to be performed at three or more levels, a posterolateral fusion should be carried out in patients less than 65 years of age.

In 1989, Johnsson et al. (57) reported the follow-up of 61 patients treated for spinal stenosis by decompression without fusion. Unlike the study of Shenkin and Hash (56), this study found a direct correlation between both preoperative and postoperative slipping and unsatisfactory outcome; postoperative slipping was encountered in 26 patients, but 8 (31%) had good results. Thirty-five patients had no slippage, and good results were obtained in 28 (80%). Postoperative slipping was more than twice as frequent in the poor-result group as in the good-result group. The postoperative results were not influenced by the progression of spondylolisthesis.

The literature supports decompression alone in the treatment of spinal stenosis if preoperative flexion-extension films do not reveal movement at regions other than a preexisting spondylolisthesis (30, 57) and if decompression is to be limited to two levels (56). Whenever preexisting hypermobility is present or decompression is required at three or more levels, a simultaneous posterolateral intertransverse fusion is warranted. In addition, there is some indication that sparing of the lateral half of the facet joints, when possible, leads to superior results (42); whenever possible, resection should spare the lateral half of the apophyseal joints.

## Spinal Instrumentation

Spinal instrumentation is in a state of rapid evolution. No one system is ideal, and the surgeon must utilize instrumentation appropriate to the disease process being treated.

**Anterior Stabilization.** Anterior spinal instrumentation, such as the Zielke, Kostuik-Harrington, and Kaneda systems, may be useful for secondary fusion procedures in complex cases but plays little role in routine degenerative lumbar spine disease. In addition, some of the older systems, such as Harrington rods, and distraction systems, such as Knodt rods, when extended to the lower lumbar spine or sacrum, can produce loss of lumbar lordosis. This painful, debilitating iatrogenic condition has been referred to as the "flat-back" syndrome (58). The indications for such systems in the lumbosacral region are quite limited (Fig. 1).

**Posterior Stabilization.** Because degenerative diseases of the lumbar spine often require posterior decompression, the instrumentation most suited for lumbar stabilization must be applied posteriorly. Among the systems in current clinical use, some are contraindicated in the lumbar spine because of loss of lumbar lordosis (Knodt rods, Harrington rods, Luque rods and rectangle, Roy-Camille). Implants that incorporate pedicular screw system include Cotrel-Dubousset (CD), Moss (Harms), AO-DCP, VSP (Steffee), Vermont spinal fixator (Krag), Edwards, Long Beach (Wiltse), and Milwaukee (Winter).

The three goals of spinal instrumentation are to correct deformity, to increase the rate of fusion, and to allow for early mobilization.

Instability of the lumbosacral junction due to degenerative disease is most commonly treated with spinal instrumentation, but primary fusion of one or two segments, whether or not limited decompression is performed, can be accomplished without the use of internal fixation and with an acceptable low rate of pseudoarthrosis (59). Spinal instrumentation is indicated when decompression has required sacrifice of greater than 50% of the facet joints or significant removal of the anulus at any level, in any decompression involving secondary procedures for failed back surgery, in any case where preoperative flexion-extension X-rays reveal instability, and for reduction of grade III or IV spondylolisthesis.

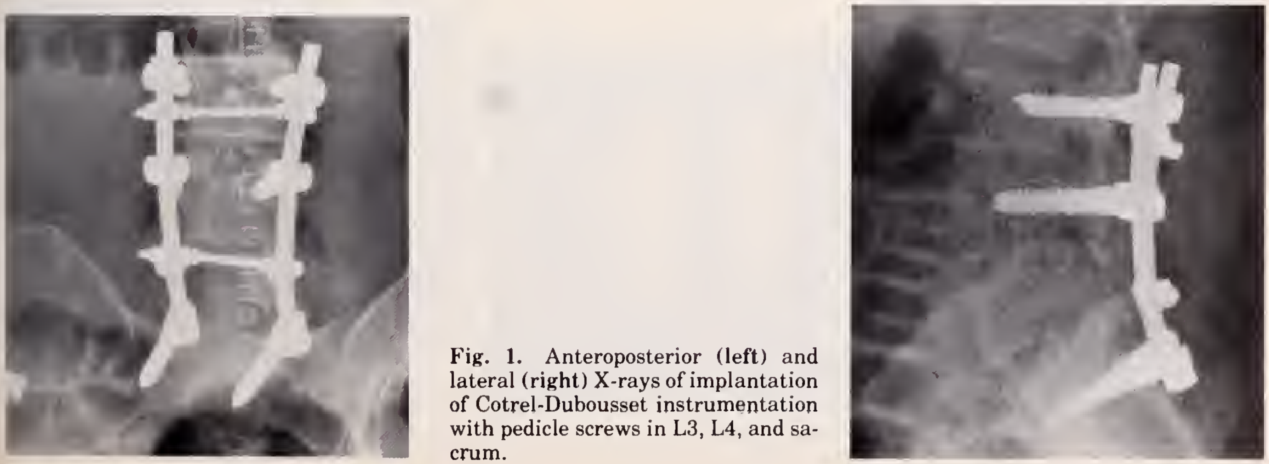


Fig. 1. Anteroposterior (left) and lateral (right) X-rays of implantation of Cotrel-Dubousset instrumentation with pedicle screws in L3, L4, and sacrum.

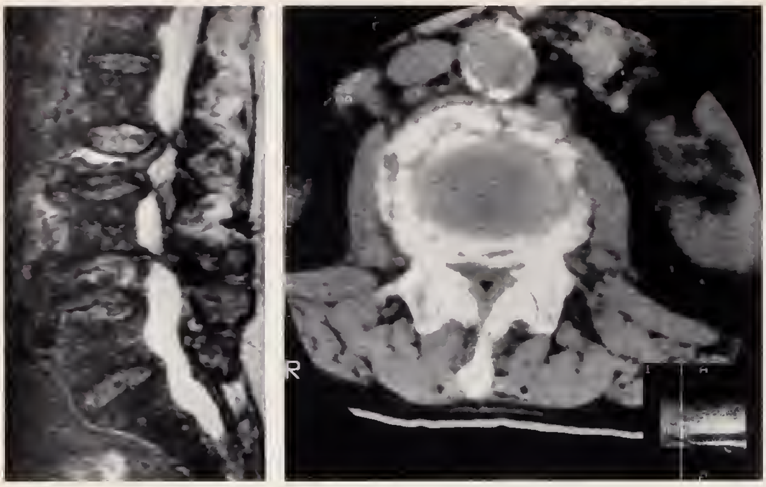


Fig. 2. Lumbar spine. **Left** Sagittal magnetic resonance imaging revealing severe multilevel stenosis between L2 and L4. **Right** Axial computed tomography. Note stenosis and collapse of vertebral body at L2 level.

When the decision for fusion is made preoperatively, as in segmental instability or a grade I or II spondylolisthesis without signs of nerve root compression, the Luque rectangle, Cotrel-Dubousset instrumentation, or facet screws may be appropriate since the posterior elements are available for fixation (5, 8, 13, 14, 59, 60).

However, the decision to fuse often occurs during surgery after a radical decompression where no posterior elements are available; in such a situation, pedicle screw fixation lends itself nicely to the repair (Fig. 2). The advantages of pedicle implants are (a) the pedicle is the strongest part of the vertebra; (b) no intact posterior elements are required; (c) three-column stability is provided; (d) shorter segments can be instrumented; and (e) the normal lumbar lordosis is preserved. The disadvantages of transpedicular devices are (a) the cantilever mechanism, which has a 10% incidence of screw fracture, (b) a 4%–10% propensity for neurologic damage, and (c) if plates are used with the screws for fixation, loss of a larger area for potential bony fusion.

With the introduction of posterior segmental instrumentation via pedicle screws, sublaminar wires, or the Cotrel-Dubousset system, much higher fusion rates are now obtained. Multilevel fusions have decreased the pseudoarthrosis rates from 20% without instrumentation to 5% with instrumentation (59). Reduction of deformities such as spondylolisthesis has been greatly enhanced by Steffee plates (Fig. 3). In addition, the need to immobilize patients until bony fusion has provided spinal stability has been eliminated.

## Summary

Spinal instrumentation to correct instability of the lumbar spine is in the midst of great

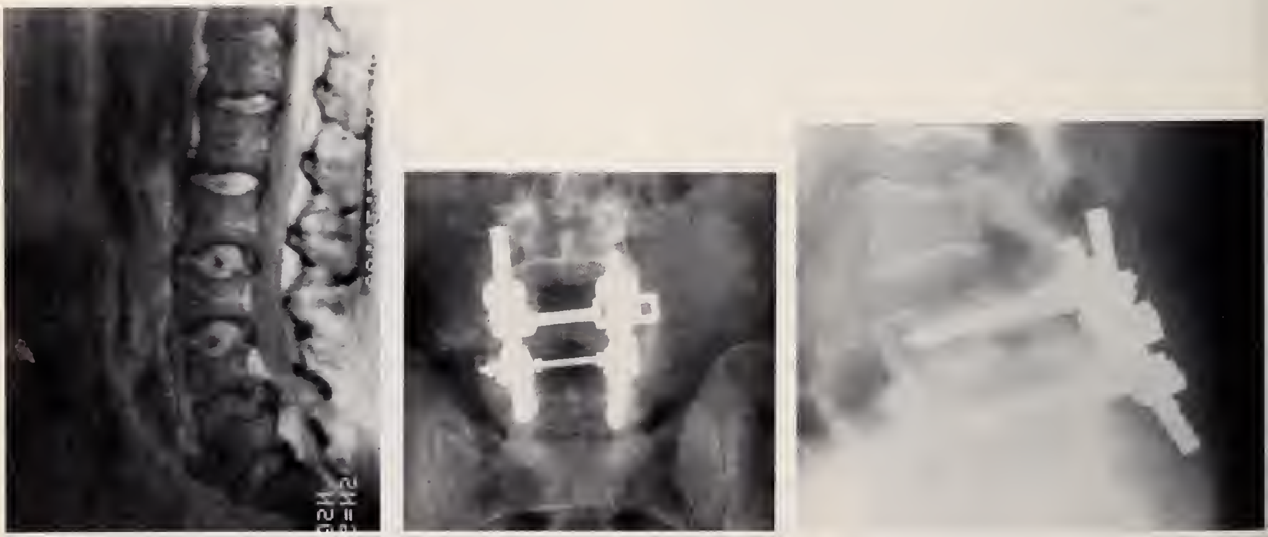


Fig. 3. **Left** Sagittal magnetic resonance imaging of lumbar spine, showing grade II spondylolisthesis at L4–5. Anteroposterior (middle) and lateral (right) X-rays. Note reduction of spondylolisthesis by Cotrel-Dubousset pedicle screws in L4 and L5.

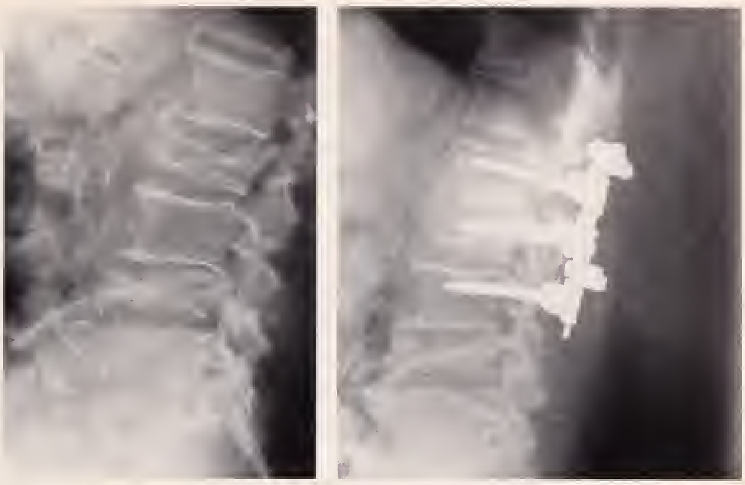


**Fig. 2 continued. Left** Lateral X-ray demonstrating severe degenerative changes and osteoporosis. Note collapse of L2 and L4 vertebral bodies. **Right** Lateral postoperative X-ray. Note stabilization with Moss pedicle screws from L1 to L3.

changes. Our review of the literature, while extensive, is not comprehensive. Only general guidelines for stabilizing hypermobility due to degenerative disease are provided.

## References

1. Hadra BE. Wiring of the spinous processes in Pott's disease. Trans Am Orthop Assoc 1891; 4:206–210.
2. Lange F. Support of the spondylitic spine by means of buried steel bars attached to the vertebrae. Am J Orthop Surg 1910; 8:344–361.
3. Albee FH. Transplantation of the portion of the tibia into the spine for Pott's disease. JAMA 1911; 57:885–886.
4. Hibbs RH. An operation for progressive spinal deformities. NY J Med 1911; 93:1013–1016.
5. King D. Internal fixation for lumbo-sacral fusion. Am J Surg 1944; 66:357–361.
6. Harrington PR. Treatment of scoliosis. Correction and internal fixation by spine instrumentation. J Bone Joint Surg 1962; 44A:591–610.
7. Bradford DS. Instrumentation of the lumbar spine. An overview. Clin Orthop Rel Res 1986; 203:209–218.
8. White AA, Panjabi MM. Clinical biomechanics of the spine. 2nd ed. Philadelphia: JB Lippincott, 1990:342–362.
9. Kostuik JP, Errico TJ, Gleason TF. Techniques of internal fixation for degenerative conditions of the lumbar spine. Clin Orthop Rel Res 1986; 203:219–231.
10. Thalgott JS, LaRocca H, Aebi M, et al. Reconstruction of the lumbar spine using AO DCP plate internal fixation. Spine 1989; 14:91–95.
11. Gurr KR, McAfee PC. Cotrel-Dubousset instrumentation in adults. A preliminary report. Spine 1988; 13:510–520.
12. Horowitch A, Peek RD, Thomas JC, et al. The Wiltse pedicle screw fixation system: early clinical results. Spine 1989; 14:461–467.
13. Kostuik JP. Anterior fixation for burst fractures of the thoracic and lumbar spine with or without neurological involvement. Spine 1988; 13:286–293.
14. Cotrel Y, Dubousset J, Guillaumat M. New universal instrumentation in spinal surgery. Clin Orthop Rel Res 1982; 227:110–123.
15. Jacobs RR, Montesano PX, Jackson RP. Enhancement of lumbar spine fusion by use of translaminar facet joint screws. Spine 1989; 14:12–15.
16. Zuckerman J, Hsu K, White A, et al. Early results of spinal fusion using variable spine plating system. Spine 1988; 13:570–579.
17. Whitecloud TS, Butler JC, Cohen JL, et al. Complications with the variable spinal plating system. Spine 1989; 14:472–476.
18. Denis F. Spinal instability as defined by the three-column spine concept in acute spinal trauma. Clin Orthop Rel Res 1984; 189:65–76.
19. Mooney V. The syndromes of low back disease. Orthop Clin N Amer 1983; 14:505–515.
20. O'Brien JP. The role of fusion for chronic low back pain. Orthop Clin N Amer 1983; 14:639–647.
21. Pope MH, Panjabi M. Biomechanical definitions of spinal instability. Spine 1985; 10:255–256.
22. White AA, Panjabi MM. The basic kinematics of the human spine. Spine 1978; 3:12–20.
23. Posner I, White AA, Edwards WT, et al. A biomechanical analysis of the clinical stability of the lumbar and lumbosacral spine. Spine 1982; 7:374–389.
24. Kirkaldy-Willis WH, Farfan HF. Instability of the lumbar spine. Clin Orthop Rel Res 1982; 165:110–123.
25. Tibrewal SB, Pearcy MJ, Portek I, et al. A prospective study of lumbar spinal movements before and after discectomy using biplanar radiography. Spine 1985; 10:455–460.
26. Paris SV. Physical signs of instability. Spine 1985; 10:277–279.
27. Prolo DJ. Toward uniformity in evaluating results of lumbar spine operations. Spine 1986; 11:601–606.
28. Dupuis PR, Yong-Hing K, Cassidy JD, et al. Radiologic diagnosis of degenerative lumbar spinal instability. Spine 1985; 10:262–276.
29. Nachemson AL. The lumbar spine, an orthopaedic challenge. Spine 1976; 1:59–71.
30. Selby DK. When to operate and what to operate on. Orthop Clin N Amer 1983; 14:577–587.
31. Nachemson AL. The role of spine fusion. Spine 1981; 6:306–307.
32. Taillard WF. Etiology of spondylolisthesis. Clin Orthop Rel Res 1976; 117:30–39.
33. Mooney V. The role of spine fusion. Spine 1981; 6:304–305.
34. Esses SI, Botsford DJ, Kostuik JP. The role of external spinal skeletal fixation in the assessment of low back disorders. Spine 1989; 14:594–601.
35. Olerud S, Sjöstrom M, Karlström G, et al. Spontaneous

effect of increased stability of the lower lumbar spine in cases of severe chronic back pain. Clin Orthop Rel Res 1986; 203:67–74.

36. Sypert GW. Low back pain disorders: lumbar fusion. Clin Neurosurg 1986; 33:457–483.
37. Nachemson AL. Lumbar spine instability. A critical update and symposium summary. Spine 1985; 10:290–291.
38. Mixter WJ, Barr JS. Rupture of the intervertebral disc with involvement of the spinal canal. N Engl J Med 1934; 211:210–215.
39. Wright A. The role of spine fusion. Spine 1981; 6:292.
40. Epstein JA. The role of spine fusion. Spine 1981; 6:281–283.
41. Hazlett JW, Kinnard P. Lumbar apophyseal process excision and spinal instability. Spine 1982; 7:171–176.
42. Panjabi MM. Degeneration, injury and spinal instability. In: Weinstein JN, Wiesel SW, eds. The lumbar spine. Philadelphia: WB Saunders, 1990:188–194.
43. Nachlas IW. End result study of the treatment of herniated nucleus pulposus by excision with fusion and without fusion. J Bone Joint Surg 1952; 34A:981–988.
44. Frymoyer JW, Hanley E, Howe J, et al. Disc excision and spine fusion in the management of lumbar disc disease: a minimum ten-year follow-up. Spine 1978; 3:1–6.
45. White AH, Von Rogov P, Zuckerman J, et al. Lumbar laminectomy for herniated disc: a prospective controlled comparison with internal fixation fusion. Spine 1987; 12:305–307.
46. Taillard W. Le spondylolisthesis chez l'enfant et l'adolescent. Etude de 50 cas. Acta Orthop Scand 1954; 24:115–144.
47. Wiltse LL, Newman PH, Macnab I. Classification of spondylolysis and spondylolisthesis. Clin Orthop Rel Res 1976; 117:23–29.
48. Newman PH. Surgical treatment for spondylolisthesis in the adult. Clin Orthop Rel Res 1976; 117:106–111.
49. Gill GG, Manning JG, White HL. Surgical treatment of spondylolisthesis without a spine fusion. J Bone Joint Surg 1955; 37A:493–520.
50. Gill GG, White HL. Surgical treatment of spondylolisthesis without spine fusion. Acta Orthop Scand 1965; 85(Suppl):4–99.
51. Amuso SJ, Neff RS, Coulson DB, et al. The surgical treatment of spondylolisthesis by posterior element resection. J Bone Joint Surg 1970; 52A:529–536.
52. Österman K, Lindholm TS, Laurent LE. Late results of removal of the loose posterior element (Gill's operation) in the treatment of lytic lumbar spondylolisthesis. Clin Orthop Rel Res 1976; 117:121–128.
53. Rosenberg NJ. Degenerative spondylolisthesis. Surgical treatment. Clin Orthop Rel Res 1976; 117:112–120.
54. Feffer HL, Wiesel SW, Cuckler JM, et al. Degenerative spondylolisthesis: to fuse or not to fuse? Spine 1985; 10:287–291.
55. Herron LD, Trippi AC. L4–5 degenerative spondylolisthesis. The results of treatment by decompressive laminectomy without fusion. Spine 1989; 14:534–538.
56. Shenkin HA, Hash CJ. Spondylolisthesis after multiple bilateral laminectomies and facetectomies for lumbar spondylosis. J Neurosurg 1979; 50:45–47.
57. Johnsson KE, Redlund-Johnell I, Udén A, et al. Pre-operative and post-operative instability in lumbar spinal stenosis. Spine 1989; 14:591–593.
58. LaGrone MO. Loss of lumbar lordosis. A complication of spinal fusion for scoliosis. Orthop Clin N Amer 1988; 19:383–393.
59. Frederickson BE, Yuan HA. Spinal instrumentation: internal fixation. In: Weinstein JN, Wiesel SW, eds. The lumbar spine. Philadelphia: WB Saunders, 1990:941–945.
60. Krag MH. Spinal instrumentation: biomechanics of transpedicle spinal fixation. In: Weinstein JN, Wiesel SW, eds. The lumbar spine. Philadelphia: WB Saunders, 1990:945–955.

# Percutaneous Radiofrequency Rhizotomy of the Lumbar Facets: Ten Years' Experience

Martin H. Savitz, M.D.

## Abstract

Over a 10-year period, 142 percutaneous radiofrequency facet rhizotomies were performed at a suburban community hospital on patients with incapacitating low back pain after various modalities of conservative treatment had failed. Neuroradiologic studies were negative or failed to reveal a herniated disk in all patients. The initial series reported in 1986 included 56 patients who had undergone previous spinal surgery and 59 patients who had lumbosacral radiculitis. An additional 27 patients with lumbosacral radiculitis have been treated. The combined use of fine thermocouple electrodes, image intensifier, nerve stimulator, and radiofrequency lesion generator provided a simple and safe technique for the relief of pain. Improvement, assessed by reduced medication requirement and increased employment, occurred in 60% of the patients.

After Rees (1, 2) described percutaneous rhizolysis of the zygapophyseal joints for relief of intractable low back pain, King (3) suggested that the operation was merely a myofasciotomy rather than a neurectomy. Shealy (4, 5) performed percutaneous radiofrequency rhizotomies of the articular nerves of Luschka with fluoroscopy. Further clinical investigation of facet denervation (6, 7) adapted the technique to practical use in a community hospital.

### Methods and Materials

There were 79 women and 58 men in the 10-year study; the ages ranged from 18 to 72 years old. Each patient was evaluated by computed tomography (CT) scan and myelography or magnetic resonance imaging (MRI). In all the patients, the neuroradiologic studies were negative or failed to confirm herniation of the nucleus pulposus. None of the patients had significant narrowing of the vertebral canal or spinal instability requiring surgery. Every patient had over an extended period of time been refractory to management by conservative measures, including trials of bed rest, traction, back brace, low back exercises, antiinflammatory agents, analgesics, muscle relaxants, physiotherapy, and transcutaneous electroneural stimulators.

The operative technique employing a Radionics model RFG 3AU was described in detail in a report published in 1986 (8). Although the C-arm readily provided both posteroanterior and lateral views, with continued experience only the posteroanterior image was needed to properly position the no. 14 gauge needles as guides for the thermocoagulating electrode (Fig.). The X-ray landmarks, as described by Fox and Rizzoli (9), were the lateral margin of each joint at the junction of the superior one-third and the inferior two-thirds. The procedure was done with the patient under neuroleptanalgesia so that he or she could respond to continuing testing. Only local anesthesia was employed in the initial series. Control of the temperature over the articular nerve roots permitted selective denervation of the pain conduction fibers. Verbal testing ascer-

---

From the Department of Surgery, Section of Neurosurgery, Nyack Hospital, Nyack, NY. Address reprint requests to Martin H. Savitz, M.D., 55 Old Turnpike Road, Nanuet, NY 10954.

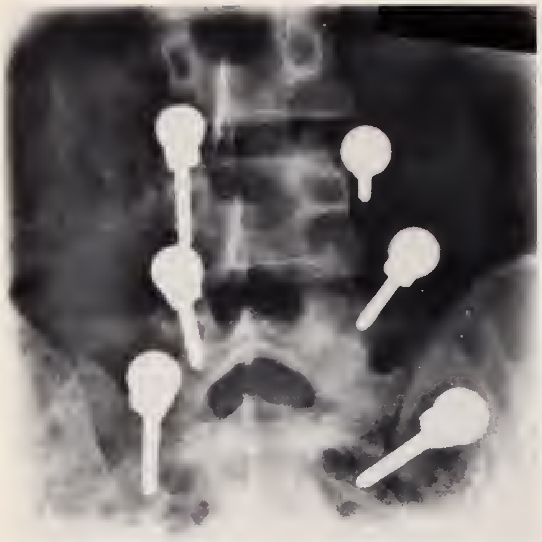


Fig. Posteroanterior view of no. 14 gauge needles employed as guides for thermocoagulating electrode at L3–4, L4–5, and L5–S1.

tained that the electrode was not too close to the anterior primary ramus and often reproduced the initial pattern of pain.

## Results

Sixty percent of the patients were able to reduce their pain medication requirement, or return to work, or both. When secondary gain was considered, the success rate was 50%, versus 70% for patients not involved in compensation or liability cases. The procedure was more beneficial (68%) in patients not previously operated on; far fewer patients (54%) with prior laminectomy improved. Five patients required repeat facet rhizotomy after more than six months of relief.

## Discussion

Evidence has been presented that mechanical derangement of the zygapophyseal joints can produce symptoms of sciatica (9–12). During preliminary stimulation of the nerves of Luschka, most of the patients described referred pain radiating down the posterior thigh. Lora and Long (6) attempted denervation of the articular facets with radiofrequency coagulation in over 200 patients with intractable pain at a university hospital. All of the patients had obtained temporary relief from local anesthetic blockade of the zygapophyseal joints. As reported by Shealy (4, 5), the success rate was highest in patients not previously operated on and fell precipitously in patients who had undergone laminectomy and spinal fusion. The experience at a community hospital (8) with percutaneous rhizotomy of the lumbar facets compares favorably with the previously published results at a pain rehabilitation center (4, 5) and a university hospital (6).

## Summary

In selected patients with debilitating low back pain, percutaneous radiofrequency facet rhizotomy under neuroleptanalgesia continues to offer treatment for patients previously considered untreatable. Although the success rate is about 60%, the morbidity remains quite low. The same-day procedure can readily be performed at small community hospitals.

### References

1. Rees WS. Multiple bilateral subcutaneous rhizolysis of segmental nerves in the treatment of the intervertebral disc syndrome. Ann Gen Pract 1971; 26:126–127.
2. Rees WS. Disconnective neurosurgery: multiple bilateral percutaneous rhizolysis (facet rhizotomy). In: Morley TP, ed. Current controversies in neurosurgery. Philadelphia: WB Saunders, 1976, 80–88.
3. King JS. Randomized trial of the Rees and Shealy methods for the treatment of low back pain. In: Morley TP, ed. Current controversies in neurosurgery. Philadelphia: WB Saunders, 1976, 89–94.
4. Shealy CN. Facets in back and sciatic pain: a new approach to a major pain syndrome. Minn Med 1974; 57:199–203.
5. Shealy CN. Percutaneous radiofrequency denervation of spinal facets. Treatment for chronic back pain and sciatica. J Neurosurg 1975; 43:448–451.
6. Lora J, Long D. So-called facet denervation in the management of intractable back pain. Spine 1976; 1:121–125.
7. Bogduk N, Long DM. Percutaneous lumbar medial branch neurotomy: A modification of facet denervation. Spine 1980; 5:195–200.
8. Katz SS, Savitz MH. Percutaneous radiofrequency rhizotomy of the lumbar facets. Mt Sinai J Med 1986; 53:523–525.
9. Fox JL, Rizzoli HV. Identification of radiologic coordinates for the posterior articular nerve of Luschka in the lumbar spine. Surg Neurol 1973; 1:343–346.
10. Goldthwait JE. The lumbo-sacral articulation: an explanation of many cases of "lumbago," "sciatica," and paraplegia. Boston Med Surg J 1911; 64:365–372.
11. Badgley CE. The articular facets in relation to low-back pain and sciatic radiation. J Bone Joint Surg 1941; 23:481–496.
12. Pedersen HE, Blunck CFJ, Gardner E. The anatomy of lumbosacral posterior rami and meningeal branches of the spinal nerves (sinuvertebral nerves): with an experimental study of their function. J Bone Joint Surg 1956; 38:377–391.

# The Risk of Wound Infection in Lumbar Disk Surgery

Sean I. Savitz, Lillian V. Lee, M.S., and Howard B. Goldstein, M.D.

## Abstract

A number of questions were left unanswered by the empirical success of a 15-year regimen of prophylactic antibiotics in preventing postoperative sepsis at three community hospitals. Although intraoperative cephalosporins have eliminated the morbidity of primary wound infection, the susceptibility to these agents of nosocomial flora has fallen considerably. The principles of antimicrobial prophylaxis have been established, but the debate over the importance of meticulous aseptic technique versus prophylactic antibiotics goes on. We investigated the microbiologic factors in lumbar disk surgery at Nyack Hospital over one year to study (a) potential sources of random contamination, (b) the flora of the operating room, and (c) the efficacy of various aspects of antiseptic routine.

In a prospective study of the risk of postoperative sepsis, Tenney et al. (1) documented 5 wound infections, over a 30-month period in 128 patients (4% risk) who underwent lumbar laminectomy without the use of prophylactic antibiotics. Based on the prior recovery of *Staphylococcus aureus* from infections of laminectomy incisions (4% rate), cephalosporins have been administered to patients undergoing lumbar disk surgery at Nyack Hospital for 15 years (2). Although the regimen of antimicrobial prophylaxis has been successful in preventing the morbidity of wound infections in clean neurosurgical patients, epidemiologic reports have shown that the susceptibility of *S. aureus* to cephalosporins decreased from 100% to 80% between 1979 and 1989. Of even greater significance is the fact that the resistance to cephalosporins of coagulase-negative *Staphylococcus* cultured from nosocomial infections has risen to 60%.

The purpose of the present study was to identify sources of contamination during lumbar disk surgery and to evaluate the efficacy of the antiseptic techniques in the neurosurgical operating theater. Additionally, the antibiotic sensitivity pattern of organisms cultured from the skin and incision site of patients and the skin and anterior nares of the operating team were tested in order to decide whether the specific regimen of antistaphylococcal prophylaxis should be altered.

## Materials and Methods

**Patient Population.** Nyack Hospital is in a suburban county approximately 20 miles from New York City; the area served by the hospital has a population of 200,000. The series reported on here included 30 men and 25 women, aged from 22 years to 76 years. All 55 patients underwent spinal surgery from August 1988 through August 1989. None of the patients had any infection prior to surgery.

**Antiseptic Routine.** The basic aseptic techniques in the operating theater did not vary from the routine previously described (2). The neurosurgical team scrubbed their hands and forearms with povidone-iodine soap. The patient's back was scrubbed with povidone-iodine soap prior to painting the operative site with povidone-iodine solution. Sterile adhesive plastic drapes were used in all cases before covering the operating table with sterile sheets. After the skin incision

---

Research supported in part by a grant from the New York Academy of Sciences, where the results were presented June 5, 1989. From the Microbiology Laboratory, Department of Pathology, Nyack Hospital, Nyack, NY 10960. Address reprint requests to Sean I. Savitz, Microbiology Laboratory, Department of Pathology, Nyack Hospital, Nyack, NY 10960.

was completed, the neurosurgeon and the assistant were double gloved. All instruments that came in contact with the patient had been rendered sterile by standard autoclave or gas sterilization. Copious irrigation with normal saline was performed throughout the procedure. Thorough attention to hemostasis prior to closure eliminated the need to leave any drains in the surgical wound.

**Antibiotic Prophylaxis.** Cefazolin 1 g was given intravenously as the incision was made. None of the patients gave a history of allergy to cephalosporins. No additional antibiotics were administered.

**Choice of Culture Media.** To insure optimum growth of aerobic and anaerobic organisms, blood agar, chocolate agar, and thioglycolate broth were all employed. In the microbiology laboratory, each agar plate inoculated in the operating room was streaked for isolation and incubated at 37°C in a $CO_2$ environment. The plates and broths were held for 72 hours and were checked each day for growth. If the plates were negative but the broth was positive, the broth was subcultured to blood agar, CDC anaerobic blood, and MacConkey agar. All plates were incubated at 37°C in $CO_2$ except for the CDC anaerobic blood agar plates, which were incubated under anaerobic conditions (Gas-Pak, BBL, Cockeysville, Md.) at 37°C and held for 2 days before being evaluated. Gram staining was done on all isolates. Coagulase (coagulase plasma, BBL, Cockeysville, Md.) testing, identification (ID), and antimicrobial susceptibility testing were performed for all staphylococci. For gram-negative bacilli, ID and antimicrobial susceptibility were determined by using the MicroScan Combo panel (Baxter Healthcare Corp., West Sacramento, Ca.). *Bacillus* species, alpha hemolytic streptococci, and *Neisseria* were not further subspeciated, nor was antimicrobial susceptibility tested. For the identification of anaerobes, an API 20A (Analytab Products, Plainview, NY) was inoculated according to the manufacturer's recommendation. If no growth occurred from the original specimens at the end of three days, the culture media were discarded.

**Culturing of Operating Team and Their Clothing.** Prior to scrubbing, skin scrapings were obtained with a sterile razor from the forearms of the neurosurgeon, assistant surgeon, and scrub nurse. Sterile cotton-tipped applicators were used to obtain samples from the nostrils.

The caps, masks, shirts, pants, and shoe covers available in the operating room dressing area were all sampled from freshly opened boxes. Cuttings were obtained with a sterile scissors and forceps.

**Culture of Surgical Wound.** Prior to scrubbing, skin scrapings were obtained with a sterile razor from the patient's back. During the surgical procedure, sterile cotton-tipped applicators were used to obtain samples of tissue fluid and to inoculate each medium at a single locus. The first intraoperative sample was taken 30 minutes after the prepared skin was opened, the second after X-ray equipment was brought into the operating room to confirm the level of exploration, and the third just prior to skin closure. Direct imprints of the surgeon's gloves were also made on the final agar plates. Prior to inoculation, each culture medium was exposed to the air for 30 minutes on a sterile stand set up next to the assistant surgeon.

**Informed Consent.** Each patient in the study was informed that the operating neurosurgeon was conducting clinical research about the efficacy of prophylactic antibiotics. Independent funding paid for materials and reagents; no patient incurred any charges.

**Avoiding Bias in Data.** The data were maintained independent of the operating neurosurgeon until the completion of the study so that he would not be influenced by premature conclusions. The regimen of antibiotic prophylaxis remained unchanged for the duration of the project.

## Results

No wound infections were documented in the 55 patients who underwent lumbar disk surgery during the 12 months of this investigation. Among the potential pathogens identified (Table), coagulase-negative *Staphylococcus* species were the most common bacterial isolate, and 10% were resistant to cefazolin. Of the 12 *S. aureus* isolates, 11 were sensitive to cefazolin, and there was one strain of methicillin-resistant *S. aureus*. Gram-negative organisms were found in skin cultures of 5 patients, in one incision site, and in 2 intraoperative cultures of the environment. All of the gram-negative bacteria were resistant to cefazolin except single isolates of *Proteus mirabilis* and *Escherichia coli*. *Peptostreptococcus anaerobius* was identified from the skin culture of one patient. A serial increase occurred in colonies of the same organism over the course of the operation in only one case. Bacteria resistant to cefazolin were found in 7% of the surgical wound cultures.

Other bacterial species encountered in this study probably reflect normal flora: micrococci,

TABLE

*Potential Pathogens Identified in 55 Lumbar Disk Operations at Nyack Hospital, NY, August 1988–August 1989*

| Organism | Surgical team | Patient's skin | Apparel | Surgical wound | Operating room environment |
|---|---|---|---|---|---|
| **Staphylococcus aureus** | | | | | |
| Methicillin-resistant | 0 | 1 | 0 | 0 | 0 |
| Cefazolin-susceptible | 1 | 3 | 0 | 3 | 4 |
| **Coagulase-negative Staphylococcus** | | | | | |
| Cefazolin-resistant | 0 | 6 | 0 | 3 | 0 |
| Cefazolin-sensitive | 3 | 38 | 2 | 25 | 14 |
| **Gram-negative bacteria** | | | | | |
| Cefazolin-resistant | 0 | 3 | 0 | 1 | 2 |
| Cefazolin-sensitive | 0 | 2 | 0 | 0 | 0 |
| **Peptostreptococcus anaerobius** | 0 | 1 | 0 | 0 | 0 |

Neisseriae, *Bacillus* species, diphtheroids, and alpha-hemolytic streptococci. Six incision sites and three cultures of the operating room environment revealed normal flora. Growth of bacteria in the surgical wound during the first 30 minutes after scrubbing occurred in less than 10% of cases. The patient's skin was identified as a source of potential pathogens in only 15% of the lumbar disk procedures. All intraoperative cultures were negative in 36% of the spinal operations. Based on individual biotyping of bacteria, no consistent source or pattern could be uncovered to account for the organisms in the surgical wound or ambient operating-room air.

## Discussion

Hunt (3) in 1933 cultured large amounts of bacteria from the closure of 28 abdominal incisions and considered the unidentified organisms to have fallen into the wound from the air. Ives and Hirshfeld (4) in 1938 conducted a more thorough study of the skin, surgical wound, and operating room environment. All of the wounds were extensively contaminated with bacteria, primarily staphylococci, and over 60% of the recorded wound infections were caused by *S. aureus*. The connection between postoperative sepsis and the presence of pathogenic bacteria in the air of the operating theater and the skin of the patient was considered proven until Howe (5) in 1956 found little correlation between the bacterial flora of the integument and subsequent wound infection. Byrne (6) in 1957 showed that the colony counts on blood agar plates exposed in the operating room were increased by the traffic of personnel and the amount of talking.

The bacteriologic contamination of surgical gloves was first investigated in 1939 by Devenish and Miles (7), who found that 24% were punctured or torn during surgery and postulated that infection occurred due to leakage of bacteria. Similar results were reported in 1951 by Girdlestone and Bourdillon (8), but the strains of *S. aureus* isolated from purulent material were not of the same phage types as those recovered from the hands of the surgical team. Wise et al. (9) in 1958 reported a 10% incidence of contamination from surgical gloves, but failed to find any source for the three wound infections recorded in a series of 71 operations.

Howe and Marston (10) in 1962 reported the most extensive search for sources of wound infection, including the operating room air and floor, fabrics in the hospital, skin of the operative site, nostrils of hospital personnel, gloves used during surgery, the surgical wound at closure, and blood cultures obtained from the patient immediately after surgery. In 10 cases out of 18 wound infections documented in 330 operations, seeding of bacteria occurred during surgery; either personnel or the patients themselves were the source. Breaks in surgical technique were implicated as the common pathway by which organisms entered the surgical incisions.

No previous study on the bacteriologic contamination of spinal surgery has been found in a review of the neurosurgical literature. Major differences in the aseptic technique now used and procedures in general surgery 20–50 years ago include the current use of adhesive plastic drapes, double gloving, and frequent lavage of the surgical wound with normal saline. In our study, by placing all of the agar plates as close to the neurosurgical procedure as possible, contamination of the mechanical barriers protecting the incision, rather than remote surfaces or the floor of the operating theater, was evaluated. Since the manufacturing standard for rubber gloves is two holes per hundred, double-gloving offers increased protection against random wound contamination (2).

Scrubbing and painting with povidone-iodine soap and solution eliminated methicillin-resistant *S. aureus* and anaerobic organisms present in one patient's skin and gram-negative bacteria present in five patients' skin from intraoperative cultures. Over one-third of the incisions were free of any bacterial growth. The efficacy of the aseptic technique was further demonstrated by the fact that random resistant staphylococci and gram-negative bacteria did not regularly propagate on serial cultures of the surgical wound.

At Nyack Hospital, antimicrobial susceptibility reports have documented the rise in resistance of staphylococci, gram-negative organisms, and anaerobic bacteria recovered from wound infections. Over 90% of the staphylococci cultured in our study were found to be sensitive to cefazolin. Resistant gram-negative organisms and anaerobic bacteria were potential sources of sepsis in 13% of the spinal operations at the community hospital we studied.

A number of questions were left unanswered by the empirical success of a regimen of prophylactic cefazolin in preventing wound infection in neurosurgical patients at one community hospital for 15 years (2, 8). The findings in the present investigation of spinal surgery differ considerably from all previous reports on the bacteriologic factors in clean abdominal surgery. The dissimilarity between the organisms cultured in the operating room we studied and the epidemiologic reports of nosocomial infections strongly suggests that the antiseptic routine alone was quite efficient in reducing the risk of postoperative sepsis. Even with the regular use of antimicrobial prophylaxis in clean patients, resistant strains of staphylococci have not flourished in the flora of our operating theater. The patient's skin was proven to be the origin of bacteria in 15% of the wounds, but other sources for the vast majority of random contamination could not be identified. The decision of the neurosurgical service at Nyack Hospital was to continue a single prophylactic dose of cefazolin and not to switch to a third-generation cephalosporin or add any additional antibiotic coverage.

### Acknowledgments

We thank the other members of the Microbiology Laboratory of Nyack Hospital for their excellent technical assistance.

### References

1. Tenney JH, Vlahov D, Salcman M, Ducker TB. Wide variation in risk of wound infection following clean neurosurgery. J Neurosurg 1985; 62:243–247.
2. Savitz MH, Katz SS. Prevention of primary wound infection in neurosurgical patients: a 10-year study. Neurosurgery 1986; 18:685–688.
3. Hunt EL. Some further observations upon contamination of operative wounds by air-borne bacteria. N Engl J Med 1933; 209:931–933.
4. Ives HR, Hirshfeld JW. The bacterial flora of clean surgical wounds. Ann Surg 1938; 107:607–617.
5. Howe CW. Prevention and control of postoperative wound infections owing to *Staphylococcus aureus*. N Engl J Med 1956; 255:787–794.
6. Byrne JJ, Okeke NE. Surgical wound infections. Am J Surg 1957; 94:398–401.
7. Devenish EA, Miles AA. Control of *Staphylococcus aureus* in an operating theatre. Lancet 1939; 1:1088.
8. Girdlestone GR, Bourdillon RB. Infections of "clean" surgical wounds by the surgeon and from the air. Lancet 1951; 1:807.
9. Wise RI, Sweeney FJ, Haupt GJ, Waddell MA. The environmental distribution of *Staphylococcus aureus* in an operating suite. Ann Surg 1958; 149:30–33.
10. Howe CW, Marston AT. A study on sources of postoperative staphylococcal infection. Surg Gynecol Obstet 1962; 115:266–275.

# Causes of Failure of Surgery on the Lumbar Spine: Ten-Year Follow-up

CHARLES V. BURTON, M.D.

IN June 1981, an interinstitutional orthopedic and neurosurgical study of the "failed-back-surgery syndrome" was published (1), identifying the primary factors causing such failures as:

| | |
|---|---|
| Lateral spinal stenosis | 58.5% |
| Recurrent or persistent disc herniation | 14% |
| Adhesive arachnoiditis | 11% |
| Central spinal stenosis | 10.5% |
| Epidural fibrosis | 7% |

(All other factors were less than 5% each.)

This identification of causes has made subsequent monitoring possible. In the article published here, the causative factors are reviewed and their status in 1991, a decade after the original study, is described.

The cost of surgery on the spine, particularly the cost of failure of surgery on the lower back, is high. The need for understanding and decreasing the incidence of such failures is imperative given the challenge to reduce health care costs. In 1980, Kane (2) estimated that 250,000 Americans underwent lumbar laminectomy each year. A survey of the published literature on the subject suggests that the overall poor results of laminectomy have been about 30% over the years, but have been less than 10% in modern comprehensive spine care centers. Although the number of laminectomies being performed each year appears to have progressively decreased over the past decade, this trend has been offset by a myriad of invasive procedures, including chemonucleolysis, percutaneous nucleolectomy, arthroscopic microdiscectomy, and now laser microdiscectomy.

Unlike some underdeveloped countries, in which the rate of postoperative sepsis is as high as 20%, in the United States infection is not a significant factor in producing failures of surgery. The overall incidence is estimated at 2.5%, but the incidence is less than 1% at spine centers. A significant decrease over the past decade may be attributed to the advocacy of Malis (3) for the routine use of prophylactic antibiotics.

The study of past failures of surgery on the lumbar spine provides the opportunity to potentiate our prowess in the future. Surgery is a significant and expensive endeavor. Neurosurgeons can no longer afford those failures which are preventable.

### Lateral Spinal Stenosis

Failure to adequately identify or treat lateral spinal stenosis was identified in the interinstitutional study as the primary cause of failures of surgery. The reasons are many but begin with continued confusion over nomenclature. For example, what could be more basic than the spelling of the word "disc" (disk?) or "anulus" (annulus?). These terms appeared 702 times in the 1989 International Society for Study of the Lumbar Spine abstract book, spelled as follows:

| | |
|---|---|
| disc, discs | 614 |
| disk, disks | 12 |
| annulus, annular | 72 |
| anulus, anular | 4 |

Although this tabulation indicates a clear-cut preference of society members, there has never been general agreement on usage. On the other hand, spelling of the term "lateral spinal stenosis" is well established, but understanding of it is typically lacking.

---

From the Institute for Low Back Care, Minneapolis, MN. Address reprint requests to Charles V. Burton, M.D., The Institute for Low Back Care, 2800 Chicago Ave., Minneapolis, MN 55407.

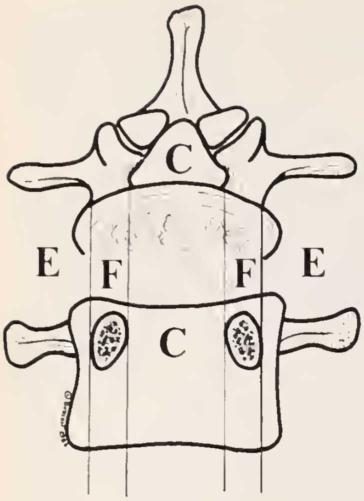
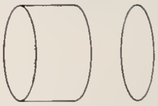


**Fig. 1. Left** Drawing of medial and lateral borders of pedicles creating three zones: central (C), foraminal (F), and extraforaminal (E). **Fig. 3. Right** Three-dimensional illustration of foramen. Entrapment may be present proximal or distal to foraminal zone.



**Central, Foraminal, and Extraforaminal Zones.** Many surgeons confuse lateral spinal with spinal recess stenosis, sometimes referred to as "subarticular stenosis." I have found the concept of central, foraminal, and extraforaminal zones helpful (Fig. 1). With enlargement of the facet joints, there is narrowing of the most lateral portion of the central zone, referred to as narrow spinal recess. When a traversing spinal nerve is compressed, spinal recess stenosis is then said to be present. This tendency may be aided by a congenitally small (trefoil-shaped) spinal canal or by an anterior osteophyte.

Fig. 2 is a three-dimensional representation of classic "up-down" lateral spinal stenosis in grade I isthmic spondylolisthesis. From the surgical standpoint, the foraminal zone is a three-dimensional rather than a two-dimensional structure (Fig. 3). It may not be evident that the area of nerve compression may extend beyond the foramen to the extraforaminal zone, in what is called the "far-out" syndrome (Fig. 4).

Appropriate deciphering of the anatomy requires quality spinal imaging. Although magnetic resonance imaging (MRI) has opened up new vistas of information, it is sometimes used in place of computed tomography (CT), which remains the single most reliable anatomic map for the surgeon. MRI is not optimal in demonstrating bone spurs and osteophytes and may miss isodense mass lesions, such as synovial cysts producing nerve compression. Adequate knowledge of the "geography of nerve entrapment" is the basis for effective surgical decompression. Primary decompression alone may not be fully adequate to restore normal nerve function.

**"Secondary" Surgical Decompression.** Some very important lessons have been learned during the past 10 years from patients with adequate decompressions who have not shown clinical improvement. Although some of these patients have permanent nerve injury, the majority of these compressed spinal nerves have not returned to optimal physiologic function. Many patients with

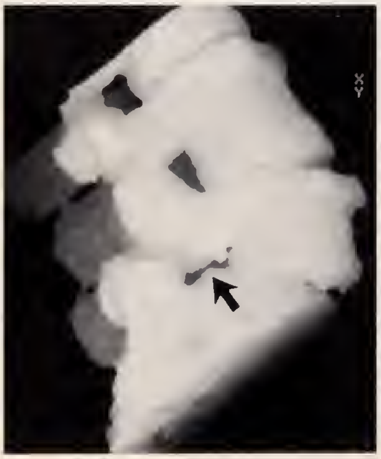
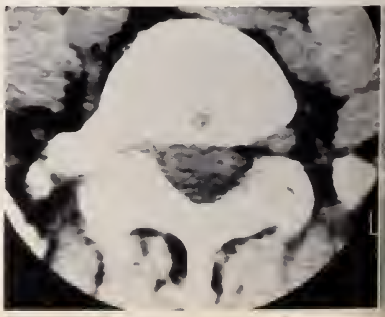


**Fig. 2. Left** Three-dimensional CT reconstruction of "up-down" lateral spinal stenosis (arrow). Pedicle of listhetic vertebra is part of stenosis. **Fig. 4. Right** Axial CT showing lateral spinal stenosis extending from foraminal zone to extraforaminal zone.



seemingly "adequate" anatomic decompressions have not achieved improvement of neurohumoral transport, axoplasmic flow, arterial supply, and venous return due to continued nerve stretch over a foraminal uncinate spur or osteophyte. Elimination of this impairment can be referred to as the "secondary" surgical decompression involved in lateral spinal stenosis surgery.

Experimental studies have shown that when a spinal nerve is compressed or stretched, the most sensitive system is venous return rather than arterial supply. This observation correlates well with the nerve swelling seen at surgery and the transformation of the distended dorsal root ganglion into a hypersensitive pain generator. "Secondary" decompressive surgery eliminates the dorsal root ganglion as a pain generator.

**Preserving Periradicular Spinal Nerve Veins.** I recently had the pleasure of meeting with Harry Crock and asked his opinion as to why apparently adequate primary and secondary decompression for LSS sometimes fails. He chided me for the obsessive manner in which neurosurgeons, with their microsurgical technique and bipolar coagulation, routinely cauterize periradicular spinal nerve veins and thus create insufficient venous return of the nerve root. I now assiduously attempt to preserve these structures and control local bleeding with other means, such as the use of thrombin-soaked patties or local thrombin/powdered Gelfoam mix.

After 10 years, lateral spinal stenosis continues to be the most common reason for failure of surgery on the lumbar spine.

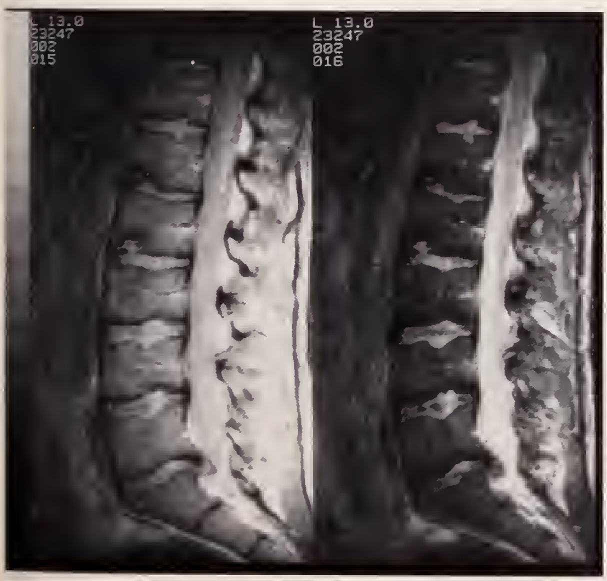


Fig. 5. Sagittal magnetic resonance imaging study of juvenile discogenic disease.

## Recurrent or Persistent Disc Herniation

In many patients with an L4–5 disc herniation, it is necessary to remove some of the L4 lamina in order to gain adequate surgical access to the disk. This relationship can be appreciated on a plain spine X-ray. The lack of evidence of a surgical defect frequently suggests the source of the problem.

Routine CT cannot differentiate between herniated disc and isodense epidural (or perineural) fibrosis. Intravenous enhancement techniques have not been very impressive and require inordinate amounts of parenteral contrast material. MRI with gadolinium has been recommended by some for the differential diagnosis. The greatest help I have found has been the double echo, spin-echo $T_2$ weighted technique. With the added advantage of being noninvasive, axial studies also permit evaluation of arachnoiditis, which is not imaged by $T_1$ gadolinium technique.

A small percentage of the population with low back pain accounts for the majority of the costs. Webster and Snook (4) recently documented that 25% of cases of low back pain accounted for 95% of the money spent. In my practice, recurrent disc herniations occur frequently in individuals (often quite young) demonstrating multilevel degenerative disc disease on MRI. The pattern of involvement tends to spare the L3 disc and typically includes internal disc disruption, multiple end-plate deformities (that is, Schmorl's nodes, limbus vertebrae), anular thickening, vertebral molding, facet degeneration, and herniated discs (Fig. 5). Most of the patients have been in the second or third decade of life, and family groups showing these changes have now been identified.

At the Center for Diagnostic Imaging in Minneapolis, Ken Heithoff and his associates surveyed 1,419 MRI scans performed during the first 3 months of 1990 and found multilevel degenerative disc changes in 9% of the patient group; males represented 75% and females 25%. For the present I have given the entity the designation "juvenile discogenic disease." There is every reason to believe that this population will prove to be a target group for recurrent disc herniation and failure of surgery.

## Adhesive Arachnoiditis

The study of foreign-body–induced lumbosacral adhesive arachnoiditis over the past 50 years, and particularly during the past decade, has been fascinating but also frightening. The medical profession has not yet succeeded in finding a benign, effective myelographic medium. Thorotrast (stabilized suspension of thorium dioxide) turned out to be a radioactive (beta emitter) contrast medium producing malignancies of the neural axis, often 20 to 30 years later. Lipiodol (combination of iodine in poppyseed oil), introduced by Sicard and Forestier in 1921, was still being recommended by Mixter (5) in 1944 because "it does no harm."

In 1944, Pantopaque (ethyl iodophenylundecylate) was introduced by Steinhausen et al. (6); based on experimental studies, they "assured" the medical profession that Pantopaque was "safe" for intrathecal injection. The conclusion, in retrospect, does not appear justified by the data. Soon after Steinhausen's (6) publication, additional trials in Sweden demonstrated toxicity, and the product was banned for human use. Surprisingly, the same level of concern was not exhibited by the remainder of the world's medical community, and not until the late 1970s did the causal relationship of iophendylate and lumbosacral adhesive arachnoiditis became clearly evident. A 1978 symposium at the International Society for Study of the Lumbar Spine reviewed the pathologic anatomy (7) (Fig. 6). Subsequently, a data base on Pantopaque toxicity was established at the Association of Consulting Toxicologists in Salt Lake City, Utah. By 1981, the role of Pantopaque as the single most significant toxic agent in the production of lumbosacral adhesive arachnoiditis was well established in the literature. Additional studies of the condition, its anatomy and physiology, and its relationship to autoimmune factors were recently reviewed and updated (8).

The disease is an insidious process that has been, and continues to be, a significant cause of failures of back surgery. During the period of maximal use, an estimated 450,000 iophendylate myelograms were performed each year in the United States. In each case, at least a mild meningeal reaction occurred. How many of these patients went on to clinically significant adhesive arachnoiditis is not known, but about 11% of total number of surgical failures were performed to treat that disease.

The condition is a cruel one. Not only is the pain typically constant and unrelenting, and thus particularly disabling, but afflicted individuals have often been considered to have functional disorders. Many physicians still maintain that iophendylate is safe and that they have never seen a case of lumbosacral adhesive arachnoiditis. Iophendylate (as Pantopaque or Myodil) continues to be used worldwide for diagnostic purposes, despite the advent of better myelographic agents, CT, and MRI scanning. As recently as 1987, the Minnesota State Neurosurgical Society passed a resolution supporting the continued use of Pantopaque as a diagnostic medium. To date, no medical, legal, or governmental ban or prohibition has ever been enforced in the United States. It is therefore not surprising that groups of patients with postmyelographic lumbosacral adhesive arachnoiditis have formed self-help groups around the world; their activity has led to the recent initiation, in both the United Kingdom and the United States, of class action legal suits against the manufacturers of iophendylate.

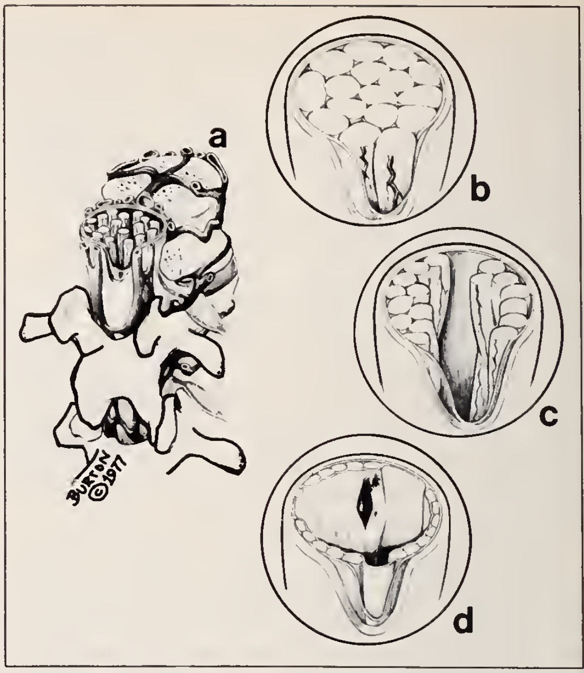


Fig. 6. Drawing of cauda equina: normal anatomy (a), adhesive arachnoiditis (b), dense fibrosis (c), and encysted globule of iophendylate (d).

## Central Spinal Stenosis

Central spinal stenosis following laminectomy continues to be a cause of failure of surgery because of confusion as to what constitutes the central spinal canal zone and the failure to adequately decompress the spinal recesses at the time of surgery.

## Epidural Fibrosis

The actual role of postsurgical scar tissue in producing failures of surgery on the lumbar spine remains unclear. To some extent, epidural and perineural fibrosis is a "normal" sequela of a spinal operation. Yet that local scarring can be a liability and significantly decreasing its presence is worthwhile. Inherent in decreasing fibrosis is also the need to reduce the volume of surgical hematoma and the promotion of a dural seal should the dura be torn. The multiplicity of present approaches to achieve these aims attests to the lack of a clear answer to the problem.

In 1978, I began to explore full-thickness autogenous fat grafts to prevent epidural fibrosis (9). Autogenous fat grafts have proven to be the most satisfactory of the various options available but have a number of drawbacks. The most satisfactory soft fat is obtained from the perigluteal area, but patients often complain of postincisional pain, and the harvesting extends the time of surgery. Too large a graft, overly dense fat, or placing fat over previously compressed dura (CSS) can produce cauda equina compression.

Despite the trial of numerous alternatives, a clearly optimal means of avoiding failures of spinal surgery due to postsurgical scarring has not yet emerged.

## Summary

A decade has passed since the primary factors in failures of surgery on the lumbar spine were identified, and the entity is still an expensive, disabling reality for too many patients. The incidence of failure has significantly decreased due to better practices as well as advances in the technology of surgical diagnostic studies and better means of conservative care. The author would like to join his colleagues in acknowledging the many contributions of Dr. Leonard Malis throughout his career toward improving the quality of patient care and helping avoid failure of back surgery in all patients.

### Acknowledgments

Particular thanks to Kenneth Heithoff, Director of the Center for Diagnostic Imaging, Minneapolis, for his review of this manuscript.

### References

1. Burton CV, Kirkaldy-Willis WH, Yong-Hing K, Heithoff KB. Causes of failure of surgery on the lumbar spine. Clin Ortho 1981; 157:191–199.
2. Kane W. Incidence of laminectomy in the U.S.A. Annual Meeting of the International Society for Study of the Lumbar Spine, New Orleans, LA, May 26, 1980.
3. Malis LI. Prevention of neurosurgical infection by intraoperative antibiotics. Neurosurgery 1979; 5:339–343.
4. Webster BS, Snook S. The cost of compensable low back pain. J Occup Med 1990; 30:13–15.
5. Mixter WJ. Spinal column and spinal cord. In: Lewis D, ed. Practice of surgery. Hagerstown, MD: WH Prior, 1944.
6. Steinhausen TB, et al. Iodinated organic compounds as contrast media for radiographic diagnoses. Radiology 1944; 43:230–234.
7. Burton CV. Lumbosacral arachnoiditis. Spine 1978; 3:24–30.
8. Burton CV. Adhesive arachnoiditis. In: Youmans JR, ed. Neurological surgery, 3rd Ed. Philadelphia: WB Saunders, 1990.
9. Burton CV. Full-thickness autogenous fat grafts in the prevention of epidural fibrosis. Contemp Neurosurg 1984; 25:1–6.

## Index to Advertisers

| | |
|---|---|
| Eli Lilly and Company | cover 2 |

This Publication
is available in Microform.

University
Microfilms
International

Please send additional information

for ____________________
(name of publication)

Name____________________

Institution____________________

Street____________________

City____________________

State__________ Zip__________

300 North Zeeb Road
Dept. P.R.
Ann Arbor, Mi. 48106

# Instructions for Authors

THE MOUNT SINAI JOURNAL OF MEDICINE, established in 1934 as *The Journal of the Mount Sinai Hospital,* is a peer-reviewed general medical journal. It is indexed or abstracted in *Index Medicus, Hospital Literature Index, International Nursing Index, Excerpta Medica, Chemical Abstracts, Biological Abstracts,* and is available on major databases. Original articles, theoretical explorations, basic science articles, reviews, reports of clinical or research experience, articles on any subject affecting the practice of medicine, short notes in any specialty, letters, and articles by medical students are welcome. The *Journal* is a participant in the Agreement on Uniform Requirements for Manuscripts Submitted to Biomedical Journals.

**Communications.** Address correspondence on manuscript submission and review to: Managing Editor, *The Mount Sinai Journal of Medicine,* Box 1094, 19 E. 98th Street, 4D, New York, NY 10029-6574. Telephone: (212) 241-6108.

**Copyright.** The publisher reserves copyright and renewal rights on all published material, and such material may not be published elsewhere without written permission of the publisher. When submitting a manuscript, include the following assignment of copyright in your cover letter: **In consideration of The Mount Sinai Journal of Medicine (MSJM) taking action and reviewing and editing my (our) submission, the author(s) undersigned hereby transfer(s), assign(s), or otherwise convey(s) all copyright ownership to The Mount Sinai Journal of Medicine in the event that such work is published by MSJM.** Each author must sign the statement.

**Manuscript Preparation.** Submit manuscripts in triplicate, including three sets of illustrations. Manuscripts must be typewritten, **double-spaced throughout** (including references, legends, and tables), on one side of 8½ × 11-in. white bond paper, with generous margins. Number pages consecutively. Begin a new page for title page, abstract, first page of text, references, each table, and legends. **Computer-generated typescripts** must be double-spaced and must be easy for editors and typesetters to read. **Computer tapes or disks:** consult managing editor.

**Title Page.** On the title page, on separate lines (double-spaced), include (1) title of article, phrased as concisely as possible; (2) name of each author, including first name and degree; (3) name and address of departments and institutions from which the work originated; (4) acknowledgments of sources of support; (5) name, place, and date of any conferences at which paper has been delivered; (6) the statement Address **reprint requests to . . . ,** giving full name and address, with zip code, of the appropriate author.

**Abstract.** Original articles should include an abstract, which should be a single paragraph not exceeding 150 words. State the purposes; basic approach or methods; main findings; principal conclusions. Emphasize new and important aspects. When an abstract is provided, a summary at the end is not necessary. In literature reviews and other articles which do not begin with an abstract, the manuscript should conclude with a short summary.

**Key Words.** Supply 3 to 10 key words or phrases at the bottom of the title page, for use in indexing the article. Use Medical Subject Headings from *Index Medicus.*

**Writing Style and Abbreviations.** Authors should aim for conciseness and clarity without repetition. Avoid medical jargon, abbreviated phrasing, and obscure or newly coined abbreviations. Define all abbreviations on first use, except those for standard, universally recognized units of measurement and statistical terms. Spell out such terms as white blood cell, hematocrit, superior vena cava.

**Units of Measurement.** Use S1 units, giving range or interval of normal values; more familiar units may be used in addition.

**Editing.** All papers will be edited to enhance conciseness and clarity without altering meaning and to insure conformity with journal style. The author is responsible for all statements in the article.

**Illustrations.** Photographs, line drawings, graphs, and charts should increase understanding of the text. Line drawings, graphs, and charts should be professionally prepared. All illustrations should be submitted in triplicate as sharp, high-contrast glossy prints. Photomicrographs must be 5⅛ in. wide. On the back of each print affix a gum label with the number of the figure (numbered consecutively in arabic numerals as cited in text), title of manuscript, name of senior author, and the word "top" with an arrow indicating top edge. Illustrations which are to appear together should be mounted accordingly as plates and should correspond to each other in size. Protect the prints by enclosing them in heavy cardboard; do not use paper clips. Photographs of patients can be published only if a copy of the permission is submitted with the photograph. Eliminate names of patients and hospital numbers from x-rays.

**Legends.** Legends should not duplicate the text. Begin legends on a separate sheet of paper, double spaced. Each legend should be numbered to match the illustration it accompanies. Illustrations mounted as plates to appear together should be accompanied by a single legend identifying the separate elements by position ("upper left," and so on) or by letters. If letters are used in the legend, a corresponding letter must appear on the print itself. Identify all abbreviations used in a figure. Indicate magnification and stain for photomicrographs.

**Tables.** Each table should be typed on a separate page, double spaced. Number tables consecutively as they appear in text, using arabic numerals ("Table 1"). Give each table a brief title (for example, "Effect of DMSO on Rats") and type it on a separate line. If abbreviations or symbols are used in the table, identify them in a footnote.

**References.** When citing references in your text, the first work cited is numbered 1, and succeeding references are numbered in sequence; enclose the citation number in parentheses and place it before any punctuation: **Cytomegalovirus (1), steroids (2), and recreational drugs (3) have all been implicated (4).** The reference list includes only works cited in the text, numbered in the order in which they are cited. Type, double spaced, following the general arrangement and punctuation style in the examples below (journal title abbreviations conform to *Index Medicus* style):

**1. Nadelman RB, Wormser GP. A clinical approach to Lyme disease. Mt Sinai J Med 1990; 57: 144–156.**

**2. International Committee of Medical Journal Editors. Uniform requirements for manuscripts submitted to biomedical journals. N Engl J Med 1991; 324: 424–428.**

**Proofs.** Proofs are sent to the corresponding author from the printer. Stylistic changes which do not alter meaning should not be made at this stage. Because of increased printing charges, the cost of excessive author's alterations beyond routine correction of typographical errors and major errors in data may be billed to the author. Proofs should be corrected and returned to the editorial office within 48 hours.

**Reprint Orders.** A reprint order form is sent to the corresponding author with proofs and is returned to the editorial office.

# THE
# MOUNT SINAI JOURNAL OF MEDICINE

## Forthcoming:

## Grand Rounds

**Bone Marrow Failure Disorders**
Blanche Alter, MD
Mount Sinai School of Medicine

**Is There A Future for Physicians in America?**
Uwe E. Reinhardt, MD
Princeton University

**Interfaces between Seizures and Affective Disorders: Evolution and Longitudinal Course**
Robert M. Post, MD, Robert L. Finding, MD, and Rene S. Kahn, MD
National Institute of Mental Health

**Interventive Gerontology: Short-Term and Long-Term Possibilities**
Gregory M. Fahy, PhD
Transplantation Laboratory
American Red Cross

**Hypertension and Left Ventricular Function**
Robert Phillips, MD
Mount Sinai School of Medicine

**Enteric Infections**
Gerald T. Keusch, MD
Tufts University School of Medicine

## Theme Issues

### MAY 1991

**Pain**
*Editor:* Allan Reed

### SEPTEMBER 1991

**Diabetes Mellitus in the 1990s**
*Editor:* Sheila Roman

### OCTOBER 1991

**Issues in Medical Ethics:**
**Surrogate Decision Making**
**Drug Use: License vs. Prohibition**
*Editors:* Daniel A. Moros and Rosamond Rhodes

### NOVEMBER 1991

**Asthma: New Concepts**
*Editor:* Alvin S. Teirstein

## Available now:

**Sexually Transmitted Diseases**
Edward J. Bottone, editor
volume 57, no. 4, September 1990
88 pages **$15**

**Inflammatory Bowel Disease Therapies for the 1990s Scientific Basis and Clinical Experience**
David B. Sachar, editor
volume 57, no. 5, October 1990
64 pages **$15**

**The Art and Science of Consultation in Medicine**
Paul L. Gilbert, editor
volume 58, no. 1, January 1991
94 pages **$15**

***Subscriptions for 1991: $60 individuals, U.S. and Canada, and $65 all libraries; $65 individuals outside U.S. and Canada.***

***To order subscriptions or copies of back issues ($15 each), please send check, payable to The Mount Sinai Journal of Medicine, to the Journal at Box 1094, One Gustave L. Levy Place, New York, NY 10029-6574; counter sales at Suite 4D, 19 E. 98th Street, New York, NY (phone: 212 241-6108; FAX: 212 831-1816).***

LEVY LIBRARY, MT. SINAI MEDICAL CEN

# THE MOUNT SINAI JOURNAL OF MEDICINE

OF THE CITY UNIVERSITY OF NEW YORK

OLUME 58 NUMBER 3 MAY 1991

## Toward the Conquest of Pain

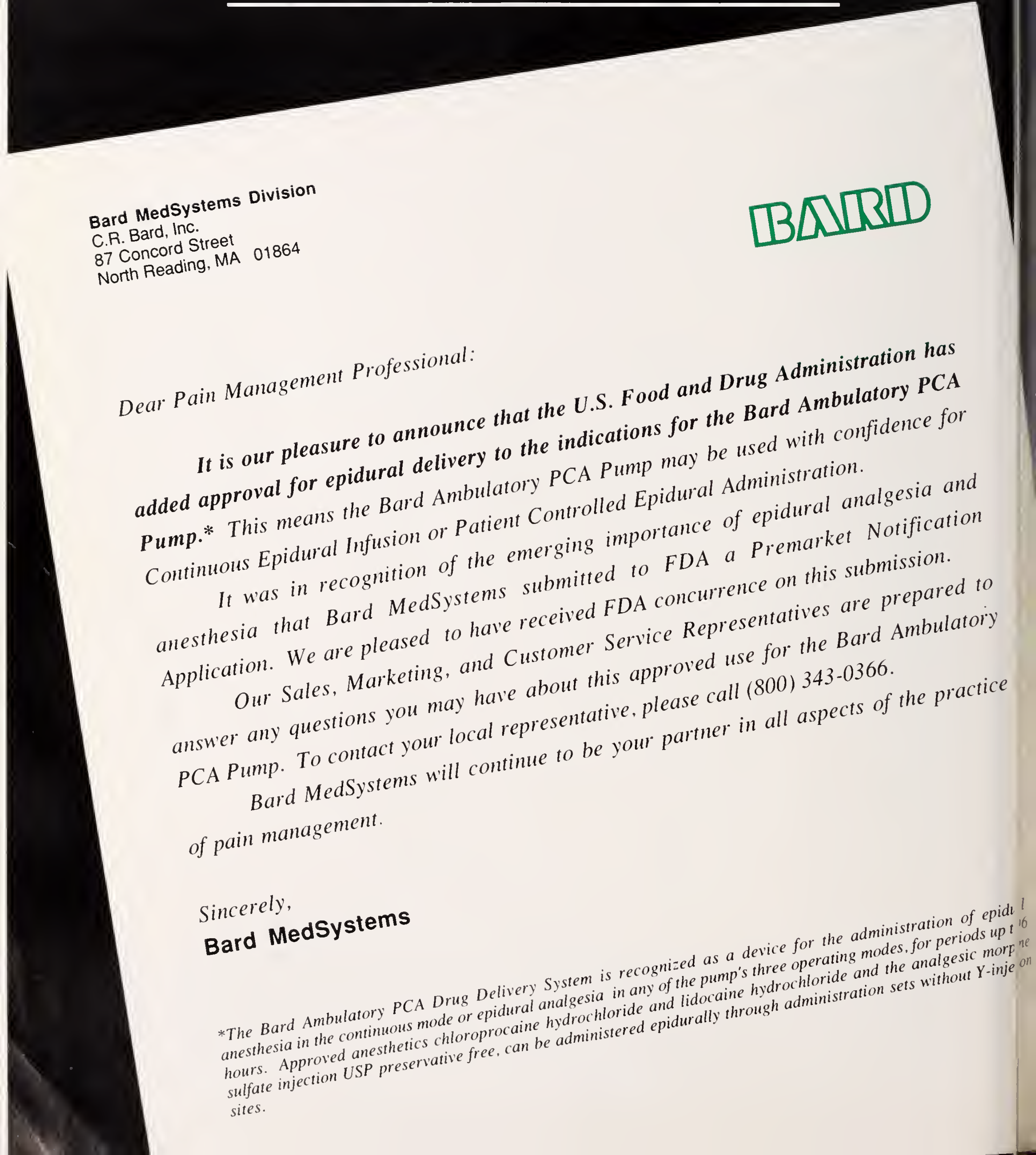
Allan P. Reed, guest editor

From the Department of Anesthesiology, Mount Sinai School of Medicine

Veterans Administration Medical Center
Bronx, New York

BETH
ISRAEL
MEDICAL
CENTER

# Important Information on EPIDURALS



**Bard MedSystems Division**
C.R. Bard, Inc.
87 Concord Street
North Reading, MA 01864

*Dear Pain Management Professional:*

***It is our pleasure to announce that the U.S. Food and Drug Administration has added approval for epidural delivery to the indications for the Bard Ambulatory PCA Pump.*** This means the Bard Ambulatory PCA Pump may be used with confidence for Continuous Epidural Infusion or Patient Controlled Epidural Administration.*

*It was in recognition of the emerging importance of epidural analgesia and anesthesia that Bard MedSystems submitted to FDA a Premarket Notification Application. We are pleased to have received FDA concurrence on this submission.*

*Our Sales, Marketing, and Customer Service Representatives are prepared to answer any questions you may have about this approved use for the Bard Ambulatory PCA Pump. To contact your local representative, please call (800) 343-0366.*

*Bard MedSystems will continue to be your partner in all aspects of the practice of pain management.*

*Sincerely,*
**Bard MedSystems**

**The Bard Ambulatory PCA Drug Delivery System is recognized as a device for the administration of epid anesthesia in the continuous mode or epidural analgesia in any of the pump's three operating modes, for periods up t hours. Approved anesthetics chloroprocaine hydrochloride and lidocaine hydrochloride and the analgesic morp sulfate injection USP preservative free, can be administered epidurally through administration sets without Y-inje sites.*

HE
OUNT SINAI JOURNAL
F MEDICINE

Volume 58
Number 3
May 1991

# CONTENTS

**Mount Sinai School of Medicine Awards and Prizes 1990/91** .......... ii

**The Mount Sinai Journal of Medicine Awards 1990** ..................... iii

## TOWARD THE CONQUEST OF PAIN

**Introduction** *Allan P. Reed* ................................................ 189

**History of Pain Concepts and Pain Therapies** *John J. Bonica* .......... 191

**Psychological Aspects of Pain** *Steven A. King* .......................... 203

**Nociceptive Transmission within the Spinal Cord** *Linda S. Sorkin* ........................................................................ 208

## CHRONIC PAIN

**The Difference between Acute and Chronic Pain** *Katherine P. Grichnik and F. Michael Ferrante* .................................... 217

**Nonnarcotic Analgesics and Tricyclic Antidepressants for the Treatment of Chronic Nonmalignant Pain** *David M. Richlin* ..... 221

**Temporomandibular Disorders and Facial Pain** *Andrew Kaplan, Daniel Buchbinder, and Suzanne McCormick* ..................... 229

**Diagnosis and Treatment of Myofascial Pain** *Hans Kraus and Andrew A. Fischer* ..................................................... 235

## ACUTE PAIN

**Update on Postoperative Pain Management** *Joel M. Kreitzer, Scott S. Reuben, and Allan P. Reed* ................................. 240

**Postoperative Pain Management in Children** *Charles M. Haberkern, Donald C. Tyler, and Elliot J. Krane* .......................... 247

**Acute Herpetic and Postherpetic Neuralgia: Clinical Features and Management** *Bradley S. Galer and Russel K. Portenoy* ........... 257

## TREATMENT

**Organization of a Multidisciplinary Pain Center** *John C. Rowlingson and Robin J. Hamill* ............................................ 267

### Index to Advertisers

**Bard MedSystems Division** **cover 2**

# Mount Sinai School of Medicine Awards and Prizes Academic Year 1990/91

Arthur H. Aufses, Sr. Prize in Surgery
**Suzette Katheryn Mikula**

Basic Sciences Achievement Award
**Brian David Golden**

Dr. Morris B. Bender Family Award in Clinical Neurology
**Sanjay Chandrasehara Rao**

Dr. Morris B. Bender Family Award in the Neurosciences
**Michael Ripps** (Class of 1993)

George James Epidemiology Award
**Craig Ross Barash**
**Howard Charles Eisenstein**
**Robert Lloyd Rosenzweig**

Doctoral Dissertation Awards
**Yiannis A. Ioannou, Ph.D.**
(Graduate School of Biological Sciences)

Harold Elster Memorial Prize for Highest Academic Standing
**Jonathan I. Seckler**

James Felt Memorial Prize for Highest Overall Performance
**Leon David Rybak**

Dr. Alan F. Guttmacher Obstetrics & Gynecology Prize
**Ilana Cass**

Saul Horowitz, Jr. Memorial Award
**Mark E. Sobel, M.D., Ph.D.** (Class of 1975)
**Idelle M. Weisman, M.D.** (Class of 1974)

Dr. Joseph R. Jagust Anesthesiology Award
**Janet R. Horn**

Dr. M. Ralph Kaufman Psychiatry Prize
**James Peter Wolberg**
**Carol F. Zale**
**Thomas M. Moriarty, Ph.D.**

Dr. Harold Lamport Biomedical Research Prize
**Daniel Jacob Goldstein**

Medical Attendings Clinical Medicine Award
**Louis Leonard Glass**

Mount Sinai Auxiliary Board Prize for Excellence in the Study of Health Care Delivery Services
**Leonard A. Shvartzman**
**Ephram Philip Weingarten**

Mount Sinai Journal of Medicine Student Prize
**David Avishai Held**
**Jonathan I. Seckler**

Florence L. Oppenheimer Surgery Prize
**Richard M. Newman**

Dr. Howard R. Rappaport Pediatric Award
**Camille Tapogna Desio**

Arthur Ross Award for Excellence
**Jennifer C. Braimon**
**Elizabeth Schneider**

Bela Schick Pediatric Society Prize
**Marie Victoria Sanford**

Nathan A. Selz Prize for Research in Cardiovascular & Renal Diseases
**Lawrence Steven Kegeles**

The Lester R. Tuchman Awards
**Karin L. Hague**
**Mary Elizabeth Schlegel**

Upjohn Clinical Excellence Award
**Alicia Diana Zalka**

Excellence-in-Teaching Awards
**Arthur I. Cederbaum, Ph.D.**
**Patrick Eggena, M.D.**
**Michael J. Klein, M.D.**
**Jeffrey T. Laitman, Ph.D.**

**Award wininers are Class of 1991 except as noted**

# The Mount Sinai Journal of Medicine Awards 1990

## Ralph Colp Prize

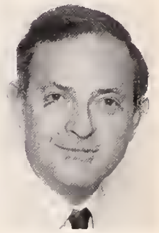
The 1990 Ralph Colp Prize for best surgical paper published in *The Mount Sinai Journal* during the calendar year went to Michael M. Todd, MD, Allan P. Reed, MD, and Ian H. Sampson, MD, for their article, "Intraoperative Management of Fluids in Neurosurgical Patients" (vol. 57, no. 6). Dr. Todd is with the Department of Anesthesiology at The University of Iowa College of Medicine. Dr. Reed and Dr. Sampson are with the Department of Anesthesiology at the Mount Sinai School of Medicine (CUNY). The award was established by colleagues and friends of Ralph Colp, distinguished surgeon and for many years chief of surgery at The Mount Sinai Hospital.

Allan P. Reed

Ian H. Sampson

## Globus Award

The 1990 Globus Award for the best paper with a clinical orientation published in this journal during the calendar year went to Albert Miller, MD, for his paper "A Simple Spirometric Clue to Asthma: Airways Obstruction Suggested by Negative or Reduced Forced Expiratory Reserve Volume despite Normal $FEV_1$-FVC Ratio" (vol. 57, no. 2). Dr. Miller is with the Pulmonary Laboratory, Division of Pulmonary Diseases, Department of Medicine, the Mount Sinai School of Medicine (CUNY). This award honors Joseph H. Globus, founder and first editor-in-chief (1934–1952) of *The Mount Sinai Journal.*



Albert Miller

## Student Prize

The 1990 Student Prize for the best student paper published in this journal has been awarded to Jonathan Seckler, BA, and David A. Held, BA, for "Patient Satisfaction with Diagnostic Cardiac Catheterization: Ambulatory vs. Inpatient Performance" (vol. 57, no. 6), written when the authors were third-year students at the Mount Sinai School of Medicine (CUNY).

Jonathan Seckler

David A. Held

# THE MOUNT SINAI JOURNAL OF MEDICINE

The Mount Sinai Journal of Medicine is published by The Mount Sinai Medical Center of New York and has the following affiliates: Beth Israel Medical Center, New York; Bronx Veterans Administration Medical Center, New York; and Elmhurst Hospital Center, New York.

**Editor-in-Chief**

Sherman Kupfer, M.D.

**Editors Emeriti**

David A. Dreiling, M.D.
Lester R. Tuchman, M.D.

**Associate Editors**

Harriet S. Gilbert, M.D.
Julius Wolf, M.D.

**Managing Editor**

Claire Sotnick

**Assistant Editors**

David H. Bechhofer, Ph.D.
Constanin A. Bona, M.D., Ph.D.
Edward J. Bottone, Ph.D.
Jurgen Brosius, Ph.D.
Steven T. Brower, M.D.
Lewis Burrows, M.D.
Joseph S. Eisenman, Ph.D.
Richard A. Frieden, M.D.
Steven Fruchtman, M.D.
Paul L. Gilbert, M.D.
James H. Godbold, Ph.D.
Richard S. Haber, M.D.
Noam Harpaz, M.D.
Dennis P. Healy, Ph.D.
Barry W. Jaffin, M.D.
Andrew S. Kaplan, D.D.S.
Samuel Kenan, M.D.
Suzanne Carter Kramer, M.Sc.
Thomas A. M. Kramer, M.D.
Mark G. Lebwohl, M.D.
Kenneth Lieberman, M.D.
Charles Lockwood, M.D.
Bernard Mehl, D.P.S.
Myron Miller, M.D.
Edward Raab, M.D.
Allan Reed, M.D.
Allan E. Rubenstein, M.D.
David B. Sachar, M.D.
Henry Sacks, M.D.
Robert Safirstein, M.D.
Ira Sanders, M.D.
Martin H. Savitz, M.D.
Clyde B. Schechter, M.D.
Phyllis Shaw, Ph.D.
George Silvay, M.D.
Barry D. Stimmel, M.D.
Nelson Stone, M.D.
Max Sung, M.D.
Rein Tideiksaar. Ph.D.
Richard P. Wedeen, M.D.
Michael F. Wesson, M.D.

**Editorial Board**

Barry Freedman, M.B.A.
Richard Gorlin, M.D.
Nathan Kase, M.D.
Panayotis G. Katsoyannis, Ph.D.
Charles K. McSherry, M.D.
Jack G. Rabinowitz, M.D.
John W. Rowe, M.D.
Alan L. Schiller, M.D.
Alan L. Silver, M.D.
Alvin S. Teirstein, M.D.

*The Mount Sinai Journal of Medicine* (ISSN No. 0027-2507; USPS 284-860) is published 6 times a year in January, March, May, September, October, and November in one indexed volume by the Committee on Medical Education and Publications (Owner), The Mount Sinai Hospital, Box 1094, One Gustave L. Levy Place, New York, NY 10029-6574. Subscription price: individuals, U.S. and Canada, $60 per year; all libraries, $65; individuals outside the U.S. and Canada, $65. Single copies, $15. New subscriptions begin with the first issue of the calendar year. Send notice of change of address 60 days before the change is effective. Second-class postage paid at New York, NY and at additional mailing offices. POSTMASTER: Send address changes to The Mount Sinai Journal of Medicine, Box 1094, One Gustave L. Levy Place, New York, NY 10029-6574. Copyright 1991 The Mount Sinai Hospital.

*Introduction*

Allan P. Reed, M.D.

# Toward the Conquest of Pain

SINCE the dawn of history and through the millennia certain fundamental aspects of life have been shared by all humankind. The agonizing and all-encompassing experience of pain is, unfortunately, one such universal.

This issue of *The Mount Sinai Journal of Medicine* attempts to address the sometimes byzantine medical questions posed by pain and its management. The pain management experts invited to contribute their expertise to this issue have been charged with the responsibility of reducing their subject matter to clinically relevant concepts that are comprehensible to the non-pain specialist.

The articles presented here divide information into two broad categories, fundamental concepts and specific disease entities.

**Fundamental Concepts.** The former group is brilliantly introduced by John J. Bonica's treatise on the historical development of pain theories and treatments. Dr. Bonica's greatest contributions to pain medicine include the establishment of the International Association for the Study of Pain; the inception of the society's journal, *Pain;* and the fostering of academic pain centers.

Three articles elucidate a sharp contrast between the science and the art of pain management. Linda S. Sorkin describes basic science principles of neuroanatomy and neurotransmitters, while David Richlin delves into the nonnarcotic pharmacology of pain management. Steven A. King attempts to quantify the more nebulous arena of psychological contributions to pain perception, as well as various methods of psychological control of pain.

The differences between acute and chronic pain are described with an emphasis on physiologic, psychological, and therapeutic implications. A multimodal approach to chronic pain evaluation is explained for those interested in establishing or altering a center devoted to chronic pain evaluation. The special high-technology options for cancer pain, implantable infusion devices, are also discussed. Reviews of the eminently successful and burgeoning areas of acute pain management for adults and children follow.

**Specific Disease Entities.** The second category of articles presents discussions of commonly encountered painful processes. Herpetic neuralgia, temporomandibular joint disease, and myofascial pain syndromes are included in this section. Although low back pain is a common problem, it has been omitted because the March 1991 issue of this journal explored the subject in depth.

**Primer in Pain Management.** This issue has been designed as a primer in pain management. Unlike major textbooks, it focuses on essential principles, rather than rehashing information readily gleaned from standard encyclopedic works or monographs. Certain disease entities that are poorly covered by the usual sources are also discussed.

**The Vexing Problem of Chronic Pain.** Over the years, pain specialists have vanquished acute pain. In 1846, an American dentist, William Morton, abolished intraoperative pain and introduced ether as a general anesthetic. Carl Koller, a German ophthalmologist, reported on the local anesthetic properties of cocaine in 1884 and thereby gave birth to regional anesthesia. Since then, innumerable refinements in technology and pharmacology have combined to eliminate or attenuate intraoperative pain, postoperative pain, and labor pain.

These successes, however, are overshadowed by the vexing problem of chronic pain. Techniques that work so well for acute pain do not reliably provide sufficient analgesia for chronic pain problems. The vicious circle of pain, decreased function, and depression continues to thwart intervention. Perennial problems such as arthritis, low back pain, and the reflex sympathetic dystrophies often elude successful therapy.

**A Fledgling Multidisciplinary Specialty.** Pain medicine is emerging as a fledgling specialty. Its standards of practice will evolve in the near future. As these events unfold, patients and

---

Allan P. Reed, M.D., guest editor of this issue, is an Assistant Professor in the Department of Anesthesiology, Mount Sinai School of Medicine.

physicians must guard against unrecognized treatments for cancer-related and non-cancer-related pain. At this time, comprehensive management of chronic pain requires a multidisciplinary approach. No single physician is adequately trained to meet the demands of such problems.

The following pages are offered as an introduction to the complexities of pain management.

# History of Pain Concepts and Pain Therapy

JOHN J. BONICA, M.D., D.SC., D.MED. SC. (HON), FFARCS (HON)

> We must all die. But that I can save him from days of torture, that is what I feel as my great and ever new privilege. Pain is a more terrible lord of mankind than even death itself.—Albert Schweitzer

THUS did Albert Schweitzer, the great humanitarian, physician, and Nobel laureate, elegantly describe in two sentences (2) the nature of pain and the obligation and privilege of the physician (and other health professionals) to relieve it. Dr. Schweitzer wrote these sentences in 1931 after nearly two decades' experience of medical practice in the African jungle.

Today, as then, proper management of pain remains one of the most important and most pressing issues of society in general and the scientific community and the health professions in particular. This importance stems from the fact that acute and chronic pain afflicts millions upon millions of persons annually, and in many patients with chronic pain and a significant percentage of those with acute pain, it is inadequately relieved. Consequently, pain is the most frequent cause of suffering and disability that seriously impairs the quality of life for millions of people throughout the world. Although accurate statistics from national and international epidemiologic studies are not available, data from a variety of sources suggest that annually in the United States and many other industrialized nations, 15% to 20% of the population have acute pain, and between 25% and 30% have chronic pain.

In most instances acute painful disorders are correctly diagnosed and effectively treated, but there is impressive evidence that all too many patients with severe postoperative and posttraumatic pain, and severe visceral pain, are not effectively relieved. In addition to needless suffering, in many patients unrelieved pain and perhaps the underlying pathophysiology cause it to progress to chronic pain. Of the patients with chronic pain, half to two-thirds are either partially or totally disabled for periods of days (for example, those with headache), some for weeks and months (for example, those with reflex sympathetic dystrophy), and some permanently (for example, those with low back pain, cancer pain, and arthritis). Because pain impairs one's ability to carry out a productive life, pain in general and chronic pain in particular is a serious economic problem as well as a major health problem.

This review is in four parts: a historical review of the concept and treatment of pain from primitive times to the middle of the present century; a summary of the theories that have evolved from the midcentury to the present time; the past and current status of pain therapy and reasons for deficiencies; and recent trends in pain research and therapy.

## Historical Perspectives

Pain is as old as humankind and perhaps even older; there is reason to believe that it is inherent in any life linked with consciousness (3). Certainly there is evidence that we have been afflicted with this evil since our beginning, for as the records of every race are examined, one finds testimonials to the omnipresence of pain. On Babylonian clay tablets, in papyri written in the days of the pyramid builders, in Persian leather documents, in inscriptions from Mycenae, on parchment rolls from Troy, down through the

---

Extracted and modified, with permission of the publisher, from Bonica JJ, History of pain concepts and therapies, chap. 1, The management of pain, 2nd ed. Philadelphia: Lea & Febiger, 1990, 2–17 (ref. 1).

From the Director Emeritus, Multidisciplinary Pain Center, University of Washington School of Medicine, Seattle, WA 98195. Address reprint requests to the author there.

ages in every civilization, in every culture are found prayers, exorcisms, and incantations bearing testimony to the prevalence of pain.

The unearthing of prehistoric human skeletons added millions of years to the recorded history of several millennia, and with it our knowledge of pain has been thrust back into the dark chasm of time, for many of these bones were indelibly stamped with signs of painful diseases, giving us evidence of how early were the beginnings of our "via dolorosa." Thus wrote the French surgeon Daetigus: "Were we to imagine ourselves suspended in timeless space over an abyss out of which the sounds of revolving earth rose to our ears, we would hear naught but an elemental roar of pain uttered as with one voice by suffering mankind" (3).

Pain has been one of the greatest factors to affect the course of human events, for scarcely anyone has escaped its throes. As classical authors related the lives of heroes, as medieval chroniclers told the legends of saints, and as biographers wrote of philosophers, artists, soldiers, inventors, scientists, and reformers, invariably one chapter of these biographies is entitled "Pain" (3). The emotional and physical consequences of persistent severe pain have been emphasized repeatedly by scientists, writers, and poets. Milton wrote in *Paradise Lost* (4):

> . . . pain is perfect miserie, the worst
> Of evils, and excessive, overturns
> All patience.

It is natural that from the beginning humankind should have engaged its energies to understand the nature of pain and attempt to control it.

## Primitive Concepts and Therapies

Apparently primitive humans had little difficulty in understanding pain associated with accidental injury but were mystified by pain caused by internal disease, or that inflicted by an arrow or spear. They treated pain from injury by rubbing (massaging) the part or exposing it to the cold water of streams or lakes, the heat of the sun, and later that of fire (5).

The cause of painful disease, or pain inflicted by a foreign object, was linked with intrusion of magic fluids, evil spirits, or pain demons into the body (6). Treatment consisted of extracting the intruding object, or making efforts to ward off, appease, or frighten away the pain demons with such ornaments as rings worn in the ears and nose, talismans, amulets, tiger claws, and similar charms. In some primitive societies, skin was tattooed with exorcist signs to keep these evil spirits outside the body. Above all, conjurations, spells, and words of might were used by the injured person, enabling the sufferer to put the pain demons to flight.

## Ancient Civilizations

### Concepts of Pain

**Ancient Egypt.** The ancient Egyptians believed painful afflictions other than wounds were caused by religious influences of their gods, or spirits of the dead, which usually arrived in darkness and entered the body through the nostril or the ear (7). The Ebers (8) and Berlin (9) papyri mention that the routes of departure of the intruding demons could be vomit, urine, the sneeze of the nose, or the sweat of the limbs. According to the Ebers papyrus (8), a widely distributed network of vessels called "metu" carried the breath of life and sensations to the heart. This was the beginning of the concept that the heart was the center of sensation (*sensorium commune*), an idea that lasted more than 2,000 years. In Assyro-Babylonian medicine, object intrusion into the body by demons or other foreign objects was the cause of pain. Thus the pain of dental caries was thought to be caused by worms burrowing into the teeth (5, 10).

**Ancient India.** The earliest concepts of pain and other medical knowledge in ancient India were attributed to the god Indra and found in the Vedas and Upanishads (11). Buddha, about 500 B.C., attributed the universality of pain in life to the frustration of desires: "Birth is attended with pain, decay is painful, disease is painful, death is painful. Union with the unpleasant is painful; painful is separation from the pleasant and any craving that is unsatisfied, that, too, is painful" (11). Though recognizing pain as a sensation, Buddhist and Hindu thought in general attached far more significance to the emotional level of the experience. As did the Egyptians, the ancient Hindus believed that pain was experienced in the heart.

**Ancient China.** Medical practice in ancient China was codified in the *Huang Ti Nei Ching Su Wen,* the Chinese canon of medicine, traced back to the time of the Yellow Emperor, Huang Ti, who lived about 2600 B.C. This work, compiled between the eighth and fifth century B.C., brings medicine from legend into the realm of history and carries it through the centuries to the practice of traditional medicine in China today (12). According to the Chinese concept, in a normal person, the two opposing unifying forces, the *Yin* (the feminine, negative, and passive force) and the *Yang* (the masculine, positive, and active force) are in balance and assist the vital energy, called the *chi,* to circulate to all parts of the body via a network of 14 channels or meridians, each connected to an important internal organ or function to which it branches. Obstruction (deficiency) or outpouring (excess) in the circulation of the *chi* causes an imbalance of the two forces and thus results in disease and pain. Acupuncture therapy at one or more of the 365 specific points located along the meridians corrects the imbalance and thus eliminates the disease and pain.

**Ancient Greece.** The ancient Greeks were intensely interested in the nature of sensory data, and the sense organs of the body found a prominent place in their physiologic speculations (see 5 for detailed review). Pythagoras (566–497 B.C.), the first great Greek thinker, who traveled widely to Egypt, Babylon, and India, apparently stimulated his disciple Alcmaeon to carry out intensive study of the senses (5). Alcmaeon, without apparent precedent, produced the idea that the brain, not the heart, was the center for sensation and reason (13). Despite the support by Anaxagoras, and subsequently by Diogenes, Democritus, and others, this view did not gain widespread acceptance in ancient Greece, due in part to the opposition of Empedocles and above all Aristotle, for whom the heart constituted the *sensorium commune.*

Plato (427–347 B.C.) believed that sensation in humans resulted from the movement of atoms communicating through the veins to the heart and liver, which were the centers for appreciation of all sensation (14, 15). He further believed that pain arose not only from peripheral stimulation, but as an emotional experience in the soul, which resided in the heart. The function of the brain in sensory processes was not clearly described; this organ was considered mainly active in elaborating the concepts derived from the sensations. He considered pain and pleasure as affections common to the whole body. Plato observed that pleasure often derived from pain relief. In the dialogue *Phaedo,* this concept is exposed by Socrates: when the chain has been taken off his leg, Socrates observes that the disappearance of pain induced pleasure (15). Plato therefore deduced that pain and pleasure, though opposite sensations, are linked together as originating from the heart as passions of the soul.

Aristotle (384–322 B.C.) elaborated on Plato's concepts on sensations and pain, mainly in *De anima* (On the soul) (16) and *Nicomachean Ethics* (17). He distinguished five senses: vision, hearing, taste, smell, and touch. For Aristotle the brain had no direct function in sensory processes; the *sensorium commune,* or center of sensory perception, was located in the heart, which he considered the most important organ of the body, that is, the center of all the fundamental life functions and the location of the soul. Pain sensation was an increased sensitivity of every sensation, but especially touch, caused by excess of vital heat. Like touch, pain arose in end organs of the flesh and was conveyed by blood to the heart. Aristotle appreciated the value of the touch-pain sense, but also emphasized that when excessively intense, it had a deleterious and indeed destructive effect (17). Like his teacher Plato, Aristotle believed that pain was felt in the heart as a "quale"—a quality or passion of the soul, a state of feeling, the experience opposite to pleasure and the epitome of unpleasantness.

Soon after Aristotle's death, Theophrastus (372–287 B.C.), his direct successor, cast doubt on his master's views, and Straton, who succeeded Theophrastus, propounded the view that the center of sensation, including pain, was in the brain (5). Later, Herophilus (335–280 B.C.) and Erasistratus (310–250 B.C.) of Alexandria, provided anatomic evidence that the brain was part of the nervous system, and nerves attached to the neuraxis were of two kinds: those for movement and those for feeling (5).

**Ancient Rome.** Among the Roman writers, Celsus (18) considered pain in relation to the phenomenon of inflammation (along with redness, swelling, and heat). Although Celsus recognized the concepts of Herophilus and Erasistratus regarding pain, particularly that of internal disease, he failed to mention the concept about the brain, the spinal cord, and motor and sensory nerves. For nearly four centuries the work of the Egyptians was lost to the Roman world until rescued by Galen (131–200 A.D.) (19).

Galen was educated in Greece and Alexandria, and then settled in Rome, where he became court physician to Marcus Aurelius. He carried out extensive studies on sensory physiology and reestablished the importance of the central and peripheral nervous system (19). He clearly established the anatomy of the cranial and spinal nerves and the sympathetic trunks. On the basis of experiments on nerves and discrete section of the spinal cord of newborn pigs, Galen elaborated a complex theory of sensation. He classified nerves as "soft," to which he attributed sensory functions; "hard," which were concerned with motor function; and a third type, related to the lowest form of sensibility—pain sensation. The center of sensibility was the brain, which was softer than any nerve and received all kinds of sensations.

Despite Galen's great contribution on the function of the nervous system in sensation, the Aristotelian concept of the five senses, and of pain as a "passion of the soul" felt in the heart, prevailed for 23 centuries. After Galen, writers of the third and fourth century summarized concepts expressed by the ancient Greek philosophers with somewhat different interpretation about what constituted the sense of touch.

### Ancient Remedies

During ancient times, the identification of the cause of pain shifted from evil spirits to the commitment of sin and the consequent punishment inflicted by an offended deity. As a result, the medicine man was replaced by the priest, servant of the gods. Along with natural remedies, the priest relied on prayers, usually made at the shrines of deities—whether the ziggurats of the Babylonians and Assyrians, the pyramids of the pharaohs, the pillared temples of the Greeks, or the teocallis of the Aztecs (3, 20). In holy ecstasy, the priests besought deities to enlighten them of the offense committed by the sufferer smitten with a painful illness, using charms and sacrifices to propitiate the immortals.

The concept of sin and punishment was also prevalent in the Judeo-Hebraic civilization, which of course believed in only one God. The same concept was adopted in the Christian ethic; the fundamental significance of the word "pain" in English derived from the Latin word *poena,* meaning punishment, and its relief through prayer. One task of Jesus Christ and his disciples was to heal the sick and to banish pain and suffering (6). Consequently, Catholicism laid great stress on and devoted much attention to the alleviation of pain by the clergy through prayer. Faith in prayer could turn every action into a remedy. The efficacy of faith in relieving pain, probably through psychotherapeutic mechanisms, has long been respected by physicians.

In addition to prayer, priests—like the medicine man of primitive people—employed natural remedies consisting mostly of herbs. The use of analgesic agents derived from plant life was prominent in all ancient cultures (6). The earliest records relate legends of the pain-relieving effects of such plants as the poppy, mandragora, hemp, and henbane. Aesculapius, the Greek god of medicine, was said to have used a potion made from herbs called *nepenthe* to produce relief of pain (3). Perhaps the first written records on the use of analgesia are those contained in ancient Babylonian clay tablets from Nippur, which date back to 2250 B.C. (8, 9). One of these describes the remedy for the pain of dental caries as a cement consisting of henbane seeds mixed with gum mastic which was applied to the cavity in the tooth (5). The Ebers papyrus, written about 1550 B.C., includes an early Egyptian pharmacopeia which contains many prescriptions for the use of opium, one of which was a remedy of divine origin prescribed by Isis for Ra's headaches (8).

The use of analgesics derived from plants and herbs was also widespread in ancient Greece. Among the earlier references to the use of pain-relieving drugs are those found in the writings of Homer, the Greek poet who lived about 800 B.C. In his *Odyssey,* he described Ulysses and his comrades being treated by Helen of Troy, daughter of Zeus, who "cast into the wine whereof they drank, a drug to lull pain and anger and bring forgetfulness to every sorrow" (21).

Celsus, in his *De medicina,* written during the first century A.D., made one of the first references to analgesic pills (18). At about the same time lived Pliny the Elder, Scribonius Largus, and Dioscorides, a Greek army surgeon in the service of Nero, all of whom wrote extensively on the preparation and use of mandragora, opium, henbane, hemp, and other drugs for the relief of pain (22, 23). Later Galen (19) wrote enthusiastically about the analgesic efficacy of opium and mandragora among the methods of pain relief.

In addition to prayer and pain-relieving drugs, the Assyrians, the Babylonians, and particularly the Egyptians used such surgical methods as trephine of the skull for headache and physical procedures such as exercise, heat, cold, and massage. Moreover, the ancient Chinese used not only acupuncture but moxibustion, massage, physical exercise, and dietary regimens to relieve painful conditions. In ancient Egypt, Greece, and Rome, electrotherapy, in the form of shocks from the Nile electric fish and the torpedo fish, was used for the treatment of neuralgia, headache, and other painful disorders (24).

### The Middle Ages, Renaissance, and Early Modern Periods

**Medieval Attitudes.** In the Middle Ages, the philosophy of Aristotle dominated, but the concept of the sensory heart was not accepted by all scientists. During this period, the center of medicine shifted to Arabia, where Avicenna (980–1038 A.D.), "the prince of physicians," proved to be the dominant figure (25). Avicenna was particularly interested in pain and the means for relieving it. In his *Canon of Medicine,* in which he codified all available medical knowledge, he distinguished five "external" senses and five "internal" senses, and located the latter in the cerebral ventricles (25). He described the etiology of 15 different types of pain due to different kinds of humoral changes and suggested such methods for their relief as exercise, heat, and massage, in addition to the use of opium and other natural drugs.

In Europe during the Middle Ages, the shift of the center of sensory perception from the heart to the brain began with the writings of Albert Magnus (26) and Mondino de' Liucci (27).

**Renaissance.** The Renaissance fostered a great scientific spirit that encouraged many remarkable advances in chemistry, physics, physiology, and anatomy, especially in the anatomy of the nervous system. During this period, Plato's and Aristotle's works were studied, with particular interest in the Greek originals and not in the Arabian translations.

During the Middle Ages and Renaissance there was virtually no advance in pain therapy. Thus Paracelsus (1490–1540) still advocated the use of opium and other natural herbs and such physical therapeutic methods as electrotherapy, massage, and exercise (6). During the latter part of the Middle Ages, the somniferant sponge, which was a sea sponge saturated with a concoction of the juices of opium, hyoscine, mandragora, and other plants, became quite popular in Europe for the relief of pain and to produce insensibility to surgical operations. Unfortunately, the effects of the sponge were unpredictable and occasionally sleep progressed to death.

**Seventeenth Century.** During the seventeenth century, additional evidence was acquired on the role of the brain in sensation, but the Aristotelian concept was still accepted by many authorities. Thus William Harvey, who in 1628 discovered the circulation of blood, still believed that the heart was the site of pain perception (4, 26). In contrast, Descartes (1596–1650), Harvey's contemporary, adhered to Galenic physiology and considered the brain the seat of sensation and motor function. In his book *L'Homme* (Man) published in 1665 (14 years after his death), Descartes described the results of his extensive anatomic and physiologic studies (28). He considered nerves as tubes that formed the marrow of the nerves and connected the proper substance of the brain with the nerves in the skin and other tissues. Sensory stimuli were transmitted to the brain by means of these threads.

**Eighteenth Century.** The eighteenth century was ushered in with the same concepts on the nature of pain and the same method for its control as had been advocated for the preceding centuries. At the beginning of the eighteenth century, the main textbooks of medicine were based on the works of Hippocrates and Aristotle, and the idea of the heart as the *sensorium commune* remained parallel to the theory that considered the brain the center of sensory perception.

**Nineteenth Century.** In the early nineteenth century physiology emerged as an experimental science; this led to the scientific study of sensation in general and pain in particular. This era was initiated in part by the publications of Bell (29, 30) and Magendie (31), who demonstrated with animal experiments that the function of the dorsal roots of spinal nerves is sensory and that of the ventral roots is motor. Impetus to the scientific study of pain was further enhanced by the writings of Johannes Müller, who in 1840 proposed the "doctrine of specific nerve energies"—that the brain received information about external objects and body structures only by way of sensory nerves, and the sensory nerves for each of the five senses carried a particular form of energy specific for each sensation (32). Müller recognized only the five classical senses—sight, hearing, smell, taste, and touch—the sense of touch incorporating for him all the qualities of experience that we derive from stimulation of the body, including the sensation of itching, pain, heat, and cold. Müller's concept was that the somatoesthetic sensations are a function of a unitary straight-through system which conveyed information from the sensory organ to the brain center responsible for sensation. These early nineteenth-century publications provoked much thought and research on sensation, particularly on pain.

During the century, significant advances were also made in pain therapy. Among the most important were the isolation of morphine from crude opium by Serturner in 1806, which was followed by the development of techniques of getting pure crystalline drugs from previously crude mixtures and allowed isolation of other opium alkaloids, including codeine in 1832. In 1828 Wohler reported the synthesis of urea, and Leroux reported the isolation of salin, which led, years later, to the introduction of salicylic acid, sodium salicylate, and acetanalid; in 1899 Dreser produced acetyl salicylic acid, which was marketed by the Bayer Company as Aspirin.

A milestone in the prevention and treatment of pain was the public demonstration of the anesthetic properties of ether by Morton in 1846, which led to the development of general anesthesia. At about this time, Rynd developed the hollow needle, and somewhat later, Wood and Pravaz developed the syringe, which permitted the injection of analgesics. This development, together with the isolation and pharmacologic studies of cocaine, eventually led to the demonstration of its local anesthetic efficacy by Karl Koller in 1884, and the subsequent widespread use of regional analgesia and anesthesia, not only for surgery but also for diagnosis and therapy of nonsurgical pain. During the century, hypnosis for surgery and for nonsurgical pain therapy came into wide use, and psychotherapeutic methods were developed.

## Pain Theories since the 1850s

In the latter part of the nineteenth century, anatomic, physiologic, and histologic studies prompted the explicit formulation of two physiologic theories of pain, the specificity theory and the intensive theory.

**The Specificity (Sensory) Theory.** The specificity or sensory theory is that pain is a specific sensation with its own sensory apparatus independent of touch and other senses. This theory, which had been suggested by Galen, Avicenna, and Descartes, and in 1853 by Loetze (33), was definitively formulated by Schiff in 1858 (34) after his analgesic experiments in animals. Schiff's theory was supported by later experiments by Funke in 1879 (35) and by classical experiments in 1882 by Blix (36) and in 1882–1885 by Goldscheider (37), who discovered separate spots for warmth, cold, and touch in the skin. A decade later von Frey (38) extended these studies to map out pain and touch spots, but he also did histologic examination of skin intended to identify specific end-organs responsible for such sensation.

**The Intensive (Summation, Pattern) Theory.** The intensive theory was implied in Aristotle's concept that pain resulted from excessive stimulation of the sense of touch; it had been suggested in midcentury by Darwin as well as Romberg, Henle, Volkmann, and Weber (33). The theory was explicitly formulated by Erb (39) in 1874, who maintained that every sensory stimulus was capable of producing pain if it reached sufficient intensity. This theory received support from Wundt (40) and later from Blix (41), Kulpe (42), Titchener (43), and especially Goldscheider (37).

In 1894 Goldscheider (44) developed the theory that stimulus intensity and central summation were the critical determinants of pain. According to this concept, pain results when the total output of cells exceeds a critical level as a result of either excessive stimulation of receptors normally fired by nonnoxious thermal or tactile stimuli, or pathologic conditions that enhance the summation of impulses produced by normally nonnoxious stimuli. Goldscheider assumed that the long delays and persistent pain observed in pathologic pain states were due to abnormally prolonged periods of summation.

**The Controversies.** Thus, by the end of the century, three conflicting concepts on the nature of pain existed. The specificity theory and the intensive theory, which were in opposition to each other, were embraced by physiologists and a few psychologists. These two theories challenged the traditional Aristotelian concept that pain was an affective quality, at this time supported by most philosophers and psychologists, including Lehmann in Germany; Bain, Bradley, Spender, and Ward in England; and Baldwin, Dewey, James, and especially H. R. Marshall in America. Marshall was the most active of the pleasure-pain theorists and wrote extensively on the subject (45).

During the decade between 1886 and 1895 proponents of each of these three theories became involved in unprecedentedly fierce controversies. In an attempt to reconcile the views of physiologists with those of philosophers and psychologists, C. A. Strong, then president of the American Psychological Association, in 1895 suggested that pain consisted of the original sensation and the psychic reaction or displeasure provoked by the sensation (46). This concept was later embraced by others, including Sherrington, who believed that pain was composed of both sensory and affective (feeling) dimensions (47).

During the first six decades of the present century, research on pain continued and the published data acquired were used in part to support either the specificity theory or the intensive theory, or a modification of them. The intense controversy between von Frey and Goldscheider continued until the late 1920s, and each rallied supporters.

**The Fourth Theory of Pain.** The controversy among the three groups of theorists continued, but by the late 1930s the specificity theory had prevailed and was universally taught (1, 48). To take into account the psychological factors that had been shown to influence pain and consolidate these with the specificity theory, Wolff, Hardy, and collaborators (48, 48a, 49) in the 1940s introduced a concept of the duality of pain modified from Strong's and called it the "fourth theory of pain." They believed that pain can be separated into two components: the sensation or perception of pain and the reactions to pain. The perception of pain, like the perception of other sensations such as temperature and touch, is a neurophysiologic process that has special structural, functional, and perceptual properties and is accomplished by means of relatively simple and primitive neural receptive and conductive mechanisms. The reaction to pain, on the other hand, is a complex physiopsychologic process involving the cognitive functions of the individual and is influenced by past experience, culture, and various psychologic factors that produce great variation in the "reaction pain threshold." This concept assumes a one-to-one relationship between the intensity of the stimulus and pain perception, and relegates the reaction to pain to a secondary role, as a straight-through, push-button consequence of sensation.

**Pattern Theories.** Goldscheider's theory prompted several modifications. In 1934, Nafe (50) suggested that all cutaneous qualities are produced by spatial and temporal patterns of nerve impulses rather than by separate modality-specific transmission routes. Based on this suggestion, and also on personal studies, two decades later Sinclair (51) and Weddell (52) proposed a peripheral pattern theory: that all fiber endings (apart from those that innervate hair cells) are alike, and thus the pattern of pain is produced by intense stimulation of nonspecific receptors. This theory ignores the physiologic evidence for a high degree of specialization of receptor fibers.

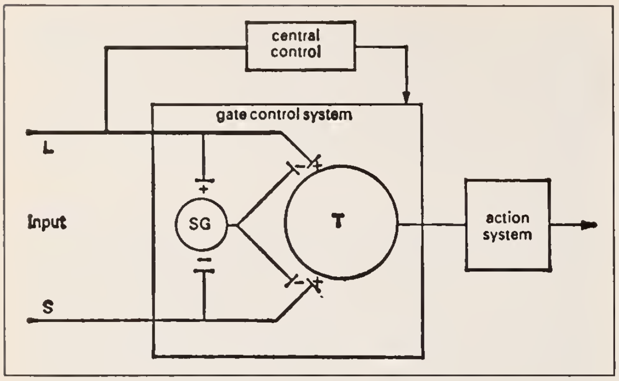
**Central Summation Theory.** In 1943, Livingston (53) proposed his own theory of central summation in support of the intensive or pattern theory. He suggested that the intensive stimulation resulting from nerve and tissue damage activates fibers that project to internuncial neuron pools in the spinal cord, creating abnormal reverberatory activity in closed self-exciting neuron loops. This prolonged, abnormal activity bombards the spinal cord transmission (T) cells which project to brain mechanisms that underlie pain perception. The abnormal internuncial activity also spreads to lateral and ventral horn cells in the spinal cord, activating the autonomic nervous system and motor system, respectively, producing increased work of the heart and skeletal-muscle spasm. These in turn produce further abnormal input, thereby creating a vicious circle. Brain activities such as fear and anxiety evoked by pain also feed into and maintain the abnormal internuncial pool activity.

**Sensory Interaction Theory.** In 1959 Noordenbos (54) proposed the sensory interaction theory, which derived from Goldscheider's original concept of two neural systems and also from proposals by Head (55), Heinbecker (56), and others, that two neural systems are involved in the transmission of pain and other sensory information: a slow system, involving the unmyelinated and thinly myelinated fibers, and a fast system, involving the large myelinated fibers. Noordenbos proposed that the small-diameter, slow-conducting somatic afferent fibers and small visceral afferents project into the cells in the dorsal horn of the spinal cord, and the summation of inputs from the small fibers produces the neural patterns transmitted to the brain to produce pain. The large-diameter, fast-acting fibers inhibit transmission of impulses from the small fibers and prevent summation. Diseases that produce a selective loss of large fibers bring about a loss of inhibition and thereby increase the probability of summation and of pain phenomena.

**Gate Control Theory.** New data acquired in the 1950s led Melzack and Wall (57) in the early 1960s to reappraise the specificity and intensive theories. They concluded that the specificity theory is strongly supported by physiologic evidence of specialization of the nervous system, but its psychological assumption that sensation is achieved via a fixed, direct-line communication from the skin to the brain in a straight-through push-button fashion is its great weakness. Similarly, the intensive theory is strongly supported by the evidence on central summation and input control but is weakened by ignoring peripheral specificity.

Wall and Melzack came to the conclusion that pain is not due to neural activity that resides exclusively in nociceptive pathways, traditionally considered specific for pain, but is a result of activity in several interacting neural systems, each with its own specialized function. As a result of these considerations, in 1965 Melzack and Wall published their own theory (57), which took into account the evidence of physiologic specialization, central summation, patterning, modulation of input, and the influence of psychological factors.

In the original Melzack–Wall theory of pain (Fig. 1), impulses evoked by peripheral stimula-



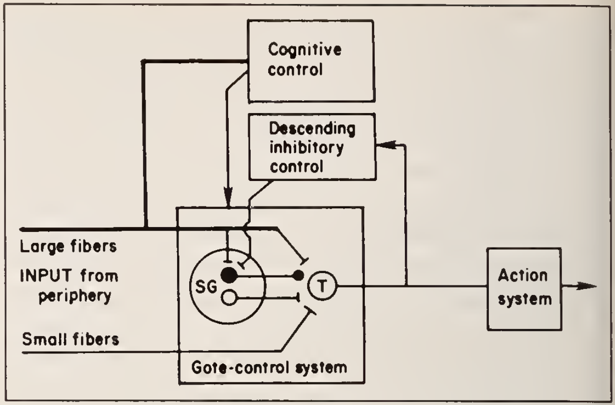


**Fig. 1. Left** Schematic diagram of gate-control theory of pain (Mark 1). Large-diameter fibers (L) and small-diameter fibers (S) project to substantia gelatinosa (SG) and first central transmission (T) cells. Inhibitory effect exerted by SG on afferent fiber terminals is increased by activity in L fibers and decreased by activity in S fibers. Central control trigger is represented by a line running from large fiber system to central control mechanisms; these mechanisms, in turn, project back to gate-control system. T cells project to action system. +, excitation; −, inhibition. *Reprinted* with permission from ref. 57, Melzack R, Wall PD. Science 1965; 150: 975 (copyright 1965 by the American Association for the Advancement of Science). **Fig. 2. Right** Gate-control theory: Mark II. New model includes excitatory (white circle) and inhibitory (black circle) links from substantia gelatinosa (SG) to transmission (T) cells as well as descending inhibitory control from brainstem systems. Round knob at end of inhibitory link implies its action may be presynaptic, postsynaptic, or both. All connections are excitatory except the inhibitory link from SG to T cells. *Adapted* from ref. 69, Melzack R, Wall PD. The challenge of pain. New York: Basic Books, 1982.

tion are transmitted to three systems: the cells in the substantia gelatinosa, the dorsal column fibers that project toward the brain, and the spinal cord transmission cells that mediate information to the brain. The theory is based on the following propositions:

- the transmission of nerve impulses from afferent fibers to the spinal cord transmission cells is modulated by a spinal gating mechanism in the dorsal horn
- the spinal gating mechanism is influenced by the relative amount of activity in large-diameter and small-diameter fibers; activity in large fibers tends to inhibit transmission (closes the gate), while small fiber activity tends to facilitate transmission (opens the gate)
- the spinal gating mechanism is influenced by nerve impulses that descend from the brain
- a specialized system of large-diameter, rapidly conducting fibers—the central control "trigger"—activates selective cognitive processes that then influence, by way of descending fibers, the modulating properties of the spinal gating mechanisms

Three years later Melzack and Casey (58) expanded the theory by taking into account subsequently acquired knowledge derived from physiologic and behavioral studies that further emphasize the motivational, affective, and cognitive aspects of the pain experience. These pertain to neural systems beyond the gate and involve interaction of the neospinothalamic and paleospinothalamic projecting systems and neocortical processes. Melzack and Casey suggested that the neospinothalamic projecting system in the brain serves to process sensory discriminative information about the location, intensity, and duration of the stimulus, while impulses that pass through the paleospinothalamic tract and paramedial ascending system activate reticular and limbic structures that provoke the powerful motivational and aversive drive and unpleasant affect that trigger the organ into action. Neocortical or higher central nervous system processes, such as the evaluation of the input in terms of past experience, exert control over both activities in both discriminative and motivational systems (Fig. 2).

## Pain Therapy Since 1900

**Development of Analgesics and Other Drugs.** During the first seven decades of the present century, the methods used to treat acute and chronic pain conceived during the nineteenth century underwent significant advances and new methods were developed. Progress was made in the development, testing, and clinical application of systemic analgesics made possible by advances in synthetic chemistry and pharmacology, including pharmacokinetics, pharmacodynamics, and controlled clinical trials. Consequently a variety of narcotic and nonnarcotic analgesics were developed and introduced into clinical practice. In addition, a heterogenous group of drugs usually labelled "adjuvants" came into use in conjunction with one or both of the other classes of drugs. During this period, and even today, drugs constitute the most frequently used method for the symptomatic relief of pain because they are readily available, inexpensive, simple to administer, and—if properly used—reasonably effective in relieving various types of pain. Unfortunately, in many instances, these drugs have been improperly prescribed or administered because of insufficient knowledge of their clinical pharmacology and misconceptions about side effects (59–61).

**Development of Regional Anesthesia and Neurosurgical Procedures.** During this period, the widespread acceptance of the specificity theory led to the development and widespread use of regional analgesia and neurosurgical operations for the treatment of nonsurgical pain. During the early part of the century, techniques were developed for the injection of local anesthetic into tissues or around virtually every major nerve and nerve plexus and into the subarachnoid and epidural spaces to produce "nerve block" or regional analgesia or anesthesia. Although these procedures were first developed for surgical anesthesia, it was not long before physicians began to use them as diagnostic, prognostic, and therapeutic aids in the management of patients with acute or chronic pain syndromes (53, 62). Moreover, the intentional destruction of peripheral nerves or sensory roots of spinal and cranial nerves by injecting alcohol, phenol, or other neurolytic agents came into wide use for the relief of severe "intractable" pain of cancer and other chronic conditions.

The specificity theory also prompted development and widespread use of neurosurgical ablative procedures intended to interrupt "pain" pathways in various parts of the peripheral and central nervous system, including such procedures as neurotomy, sensory rhizotomy, and section of the spinothalamic pathways in the upper thoracic and upper cervical region, medulla, or mesencephalon. During this period, the use of hypnosis for pain therapy underwent periods of enthusiastic acclaim and total neglect. Psychotherapy for patients with persistent pain was also used. Moreover, a variety of physical therapeutic techniques and radiation therapy were widely applied for the treatment of acute and chronic pain.

**Assessments of Pain Management.** During World War II, I became heavily involved in the treatment of military personnel with acute and chronic pain. As a result of this experience, I realized there were significant differences between acute and chronic pain and I developed and put into practice the concept that complex pain problems could be more effectively managed by a group of specialists from different disciplines working in a well-coordinated manner, each member contributing his or her specialized knowledge and skill to the common goal of making a correct diagnosis and developing the most effective therapeutic strategy. In 1953 I wrote the first comprehensive text on pain which subsequently became known as "the Bible" of the field of pain treatment. My research for the book made me very much aware that many patients with acute and chronic pain were not being effectively relieved.

Subsequently, I reassessed the status of pain research and therapy at frequent intervals, and each time I found impressive evidence that a large percentage of patients with acute postoperative pain, post-traumatic pain, myocardial infarction pain, and pain of visceral disease, and most patients with acute postburn pain, were inadequately relieved (1, 63). Moreover, many patients with nonmalignant chronic pain did not respond to the usual therapies because the therapies were improperly applied or because medical or surgical therapy was ineffective. An impressive number had been exposed to the high risks of iatrogenic complications, including drug toxicity, drug dependence, and multiple, often useless and at times mutilating surgical operations. A significant number of these patients gave up medical care and consulted quacks, who not only depleted their economic resources but, not infrequently, did them harm. Patients with cancer pain fared no better, and indeed numerous reports have indicated that many patients with advanced cancer have lived the last months of their lives in unrelieved severe pain (64).

On each reassessment of pain diagnosis and therapy, I found the reasons for these deficiencies remained the same. These many reasons were arbitrarily grouped into three major categories:

- great voids in knowledge about pain and its mechanisms due to insufficient and indeed neglect of pain research
- inadequate or improper application of the available knowledge and therapies because of lack of education of medical students and practitioners
- problems with transfer of information and communication among basic scientists and between this group and clinicians

Over the years I discussed these issues on numerous occasions and in several dozen publications in this and a number of other countries (65, 66) in an attempt to make the public, the scientific community, and the health professions aware of this serious problem. On these occasions

I advocated the need for facilities offering diagnosis and therapy for pain, including facilities of the multidisciplinary type. I also pointed out the need for an international society and a journal devoted exclusively to pain research and therapy. Unfortunately, for a quarter century, my pleadings were ignored, and pain research and therapy remained inadequate.

### Recent Advances

Fortunately, during the past two decades, several developments have taken place which, if sustained and expanded, hold the promise of helping to rectify some of these deficiencies in the management of pain. This most encouraging trend was provoked by a number of converging factors, including publication of the Melzack-Wall and the Melzack-Casey theory of pain, which stimulated basic research; increased interest in developing pain clinics and pain centers, in part provoked by my persistent drum-beating, the extensive widespread publicity about our pain program, and the organization of similar facilities by my former pain fellows in their new institution; and the founding of the International Association for the Study of Pain (IASP) in 1974 and the beginning of publication of the journal *Pain* a year later. These and other factors initiated what has become a worldwide movement in pain research and therapy.

The impressive increase in pain research carried out by scientists and clinical investigators has resulted in the acquisition of much new knowledge on the anatomy, physiology, biochemistry, and psychology of acute pain and some new information on the pathophysiologic, psychological, and environmental substrates of chronic pain (65–67). This new knowledge has changed our concept of sensory processing and sensory coding and, together with recent clinical experience, has changed our concepts and approaches to diagnosis and therapy of acute and chronic pain.

During this period, progress in the communication and transfer of information has also been impressive. The founding of the IASP and the journal *Pain* must be considered among the most important developments in pain research and therapy during the past half century. In the short period of 17 years, IASP has grown to a membership of over 4,000 persons, representing every biomedical science, clinical discipline, and health profession from 54 countries, and has fulfilled most of its purposes to an impressive degree. Through the scientific programs of the six world congresses on pain sponsored by the IASP and publication of the congress proceedings and of the journal *Pain,* IASP has done much to disseminate basic and clinical information of multidisciplinary interest. The Committee on Taxonomy has completed its task of developing definitions and a classification of various pain syndromes (68).

Progress has also been made in improving the care of patients with acute and chronic pain. The recent worldwide movement of attention to pain has prompted an increasing number of physicians and other practitioners to acquire more knowledge about pain and its treatment. Moreover, an increasing number of physicians and other health professionals devote part or all of their professional time to pain diagnosis and therapy; the number of clinics and similar facilities for pain diagnosis and treatment has increased commensurately (63).

The new knowledge acquired in recent years has prompted a revision in approaches to diagnosis and therapy of pain, including the widespread application of neurostimulation and a commensurate increase in the use of certain physical therapeutic modalities as part of pain therapy programs. Recent advances in chemistry, pharmacology, and pharmacokinetics have prompted the development of more potent nonnarcotic and narcotic agonist-antagonist analgesics; the use of other adjuvants such as psychotropic drugs for analgesia and treatment of reactive depression associated with certain chronic pain states; better prescription and administration based on recent pharmacokinetic data, including continuous-infusion and patient-controlled analgesia; and the advent and widespread use of intraspinal narcotics.

The increased surge of interest in the multidisciplinary approach to the diagnosis and therapy of chronic pain, and the hospice concept of managing patients with terminal cancer and other chronic disease, are considered by Melzack and Wall (69) and others to be the two most important advances in patient care.

### Summary

The development of the concept of pain and its treatment constitutes one of the most interesting and fascinating chapters in the history of medicine. The concept proposed by Aristotle and other ancient Greek philosophers that pain was a passion of the soul remained widely accepted for some 22 centuries, and treatment remained empirical and often ineffective. For a century after the scientific study of pain began, several theories were proposed, and these theories prompted the development of various therapeutic modalities. However, until two decades ago, pain research remained conceptually stagnant; the meager work done was not commensurate with the magnitude and clinical importance of pain. Consequently, pain treatment remained somewhat empirical and ineffective. Whatever knowledge and effective therapeutic modalities were available were not properly applied, primarily because medical students and physicians were not taught the basic principles of pain management. Fortunately, during the past two decades, significant advances have been made in our knowledge of basic mechanisms and a variety of new modalities have been introduced and old ones refined. Although we should be proud of these advances, much more needs to be done if the millions of patients with acute and chronic pain are to be managed effectively. This requires sustaining and expanding research programs; educating and training health professionals; and informing the public.

## References

1. Bonica JJ. History of pain concepts and therapies. In: Bonica JJ. The management of pain, 2nd ed. Philadelphia: Lea & Febiger, 1990, 2–17.
2. Schweitzer A. On the edge of the primeval forest. New York: Macmillan, 1931, 62.
3. Fülop-Miller R. Triumph over pain, trans. E. and C. Paul. New York: Literary Guild of America, 1938, 7–8.
4. Milton J. Paradise Lost, book 6, line 462.
5. Keele KD. Anatomies of pain. Oxford: Blackwell, 1957.
6. Tainter ML. Pain. Ann NY Acad Sci 1948; 51:3.
7. Krüsen FH. Physical medicine. Philadelphia: WB Saunders, 1941.
8. Ebers Papyros: Das hermetische Buch uber die Arzneimittel der alten Aegypter von Georg Ebers. Leipzig: Engelmann, 1875.
9. Wreszinski W. Der grosse medizinische Papyrus des Berliner Museums. Leipzig: Hinrichs, 1909.
10. Sigerist HE. A history of medicine, vol. 1. New York: Oxford University Press, 1951, 274, 349.
11. Lin Yutang. The wisdom of India. Cited in Keele KD. Anatomies of Pain. Springfield, IL: Charles C Thomas, 1957, 7.
12. Veith I. Huang, Ti Nei Ching Su Wen. Baltimore: Williams and Wilkins, 1949, 227.
13. Soury J. Le Systeme nerveux centrale. Paris: Carre et Naud, 1899, 53.
14. Plato. Timaeus, ed. RD Archer-Hind. London: Macmillan, 1888.
15. Plato. Phaedo, ed. WD Geddes. London: Macmillan, 1885.
16. Aristotelis de anima libri tres, ed. A Frider Trendelenburg. Berlin: W Weberi, 1877.
17. Aristotelis Ethica Nichomachea, ed. G Ramsauer. Leipzig: BG Teubueri, 1878.
18. Celsus AC. De medicina, vol. 15, trans WG Spencer. Cambridge, MA: Harvard University Press, 1891, 1935–1938.
19. Galen C. De usu partium, trans M Tallmadge. Ithaca, NY: Cornell University Press, 1968.
20. Castiglioni A. A history of medicine. New York: AA Knopf, 1947.
21. Homer. Iliad, trans. Edward, Earl of Derby. Philadelphia: Henry T. Coate, n.d., n.v., book 9, lines 963–967.
22. Macht DI. The history of opium and some of its preparations and alkaloids. JAMA 1959; 64:477–481.
23. Dioskurides. The Greek herbal of Dioscorides, trans. J Goodyer, 1655. London: Oxford University Press, 1934.
24. Kane K, Taub A. A history of local electrical analyses. Pain 1976; 11:125.
25. Avicenna. The canon of medicine, trans OC Gruner. London: Luzae, 1930.
26. Procacci P, Maresca M. The pain concept in western civilization: a historical review. In: Advances in the Management of Pain—Advances in Pain Research and Therapy, vol. 7. Benedetti C, CR Chapman, G Moricca, eds. New York: Raven Press, 1984; 1–13.
27. Mondino de' Liucci. Anathomia. Venice: D. Bernardini, 1538.
28. Descartes R. L'homme. Paris: Angot, 1664.
29. Bell C. Idea of a new anatomy of the brain submitted for the observations of his friends. London: Strahan & Preston, 1811.
30. Bell J, Bell C. The anatomy and physiology of the human body, 5th American ed. New York: Collins, 1827.
31. Magendie F. Experiences sur les fonctions des racines des nerfs rachidiens. J Physiol Exp 1822; 2:276–279.
32. Müller J. Handbuch der Physiologie des Menschen, vol. 2, trans. W Baly. London: Raylor and Walton, 1840, 249–250.
33. Dallenbach KM. Pain: history and present status. Am J Physiol 1939; 52:331–347.
34. Schiff M. Lehrbuch der Physiologie des Menschen I: Muskel, und Nervenphysiologie. Schavenburg: Lahr, 1858; 234:253–255.
35. Funke O. Der Tastsium und die Geneingefuhle. In: Herman L, ed. Handbuch der Physiologie der Sinnesorgane. Leipzig: Vogel 1879; 289–309.
36. Blix M. Experimentelle Beitrag zur Losung der frage über die specifische Energie der Hautnerven. Z Biol 1884; 20:141.
37. Goldscheider A. Neue Tutsachen über die Hautsinnesnerven. Arch Anat Physiol (Physiol Abth Suppl Bd) 1885; 1–110. Gesammelte Abhandlungen 1, 1898, 197).
38. Von Frey M. Beitrage zur Physiologie des Schmerzsinnes. Ber Verhadl konig sachs Ges Wiss Leipzig 1894; 46:185–196, 288–296.
39. Erb WH. Krankheiten der peripherischen cerebrospinalen Nerven. Quoted in GWA Luckey. Some recent studies of pain. Am J Psychol 1895 (1874); 7:109.
40. Wundt W. Grundzuge der Physiologischen Psychologie. Leipzig: Engelmann, 1874.
41. Blix M. Experimentelli Beitrag zur Losung der Frage über die specifische Energie der Hautnerven. Z Biol 1885; 21:160.
42. Kulpe O. Outlines of psychology, trans EB Titchener, 1905. Original Publication: Grundriss der Psychologie. Leipzig: Engelmann, 1893.
43. Titchener EB. An outline of psychology. New York: Macmillan, 1896, 65.
44. Goldscheider A. Ueber die Summation von Hautreizen. Arch Physiol 1891; 164–169. (Gesammelte Abhandlungen 1, 1898, 397–432).
45. Marshall HR. Pleasure-pain and emotion. Psychol Rev 1895; 2:57–64.
46. Strong CA. The psychology of pain. Psychol Rev 1895; 2:329.

47. Sherrington CS. Cutaneous sensations. In: Schafer EA, ed. Textbook of physiology, vol. 2. London: Pentland, 1900, 920–1001.
48. Wolff HG, Gasser HS, Hinsey JC, eds. Pain, vol. 23. Research Publications Association for Research in Nervous and Mental Disease, Baltimore: Williams and Wilkins, 1943.
48a. Wolff HG, Wolf S. Pain. Springfield, IL: Charles C Thomas, 1948.
49. Hardy JD, Wolff HG, Goodell H. Pain sensations and reactions. Baltimore: Williams & Wilkins, 1952.
50. Nafe, JP. The pressure pain and temperature senses. In: Murchison CA, ed. Handbook of general experimental psychology. Worcester, MA: Clark University Press, 1934.
51. Sinclair DC. Cutaneous sensation in the doctrine of specific nerve energy. Brain 1955; 78:584–614.
52. Weddell G. Somesthesis in chemical senses. Ann Rev Psychol 1955; 6:119–136.
53. Livingston WK. Pain mechanisms. New York: Macmillan, 1943.
54. Noordenbos W. Pain, Amsterdam: Elsevier Press, 1959.
55. Head H. Studies in neurology. London: Oxford University Press, 1920.
56. Heinbecker P, Bishop GH, O'Leary J. Pain and touch fibers in peripheral nerves. Arch Neurol Psychiat 1933; 29:771.
57. Melzack R, Wall PD. Pain mechanisms: a new theory. Science 1965; 150:971–975.
58. Melzack R, Casey KL. Sensory, motivational, and central control determinants of pain. In: Kenshalo DR, ed. The skin senses. Springfield, IL: Charles C Thomas, 1968; 423–439.
59. Marks RM, Sachar EJ. Undertreatment of medical inpatients with narcotic analgesics. Ann Int Med 1973; 78:173–181.
60. Cohen FL. Postsurgical pain relief: patients' status and nurses' medication choices. Pain 1980; 9:265.
61. Bonica JJ, Procacci P. General considerations of acute pain. In: JJ Bonica, The management of pain, 2nd ed. Philadelphia: Lea & Febiger, 1990; 159–179.
62. Bonica JJ. Regional analgesia/anesthesia. In: JJ Bonica, The management of pain, 2nd ed. Philadelphia: Lea & Febiger, 1990; 1878–1966.
63. Bonica JJ. Multidisciplinary/interdisciplinary pain programs. In: JJ Bonica, The management of pain, 2nd ed. Philadelphia: Lea & Febiger, 1990; 197–208.
64. Bonica JJ. Cancer pain. In: JJ Bonica, The management of pain, 2nd ed. Philadelphia: Lea & Febiger, 1990; 400–460.
65. Bonica JJ. Pain research and therapy: past and current status and future needs. In: Ng LKY, Bonica JJ, eds. Pain, discomfort and humanitarian care. Amsterdam: Elsevier/North Holland, 1980, 1–46.
66. Bonica JJ. IASP Presidential Address—Pain research and therapy: achievements of the past and challenges of the future. In: Bonica JJ, Lindblom U, Iggo, A, eds. Advances in pain research and therapy, vol. 5. New York: Raven Press, 1983, 1–36.
67. Bonica JJ. Anatomic, physiologic and biochemical bases of nociception and pain. In: JJ Bonica, The management of pain, 2nd ed. Philadelphia: Lea & Febiger, 1990; 28–121.
68. Merskey H. Development of a universal language of pain syndromes. In: Bonica JJ, Lindblom U, Iggo A, eds. Advances in pain research and therapy. vol. 5. New York: Raven Press, 1983; 37–52.
69. Melzack R, Wall PD. The challenge of pain. New York: Basic Books, 1982.

# Psychological Aspects of Pain

Steven A. King, M.D.

Of all the problems faced by physicians, pain is among the most pervasive and difficult to categorize and treat (1). To successfully care for the patient with pain, many preconceptions must be abandoned.

In the past, pain has frequently been perceived as having either a physical or psychological etiology. However, recent advances in the study of pain are changing this attitude. In light of research on neurotransmitters and discoveries on the ways in which the mind can influence the functioning of the body, the sensation of pain must now be viewed as reflecting the interaction between mind and body.

Furthermore, the concept that pain can be divided into two simple categories—"real," based on physical illness, or "imaginary"—is no longer viable. "Pain occurring in unicorns, griffins and jabberwockies is always imaginary pain, since these are imaginary animals; patients, on the other hand are real, and so they always have real pain" (2).

Thus, it is vital for the physician who evaluates and treats the patient with pain to investigate not only the possible underlying organic pathology, but also concomitant psychological factors.

It is the purpose of this paper to provide an overview of current knowledge on the psychological aspects of pain and psychologically based treatment approaches to pain.

## Evaluating the Patient With Pain

**Acute versus Chronic Pain.** Although much greater attention has been paid to the role of psychological factors in chronic pain than in acute pain, these factors are significant in both. Research has shown that many types of pain, ranging from traumatic to postoperative, are strongly influenced by the mental state of the patient (3).

With regard to associated psychological symptomatology, acute pain is considered to be related more to anxiety, chronic pain, to depression. The treatment approach to each is different. For example, in acute pain, allowing the patient to have at least some measure of control over his or her environment is important. In contrast, patients with chronic pain often employ pain as the primary method of control in their lives. This, in turn, reinforces the existence of the pain and therefore is ultimately detrimental. Furthermore, although antianxiety medications may be useful in the treatment of acute pain, they are generally considered to have minimal benefit in management of chronic pain and may even be contraindicated (4).

**Current Concepts of Pain.** Because the traditional views of pain have been found to have limited applicability, alternative formats have been proposed. For example, Black described the "chronic pain syndrome": the patient has "intractable, often multiple pain complaints, which are usually inappropriate to existing somatogenic problems; multiple physician contacts and many nonproductive diagnostic procedures; excessive preoccupation with the pain problem; [and] an altered behavior pattern with some of the features of depression, anxiety, and neuroticism" (5). However, others have questioned the validity of this diagnosis (6, 7).

Because pain is a subjective experience influenced by a variety of factors, the concept of "illness behavior" has been found useful (8, 9). Criteria included in illness behavior are pain perception; decisions on whether to seek treatment, and if so, from whom; the meaning of the pain to the patient; the manner in which the patient communicates about pain; and the effect

---

From the Department of Psychiatry, Maine Medical Center, Portland, Maine. Address reprint requests to Steven A. King, M.D., Director, Pain Service, Department of Psychiatry, 22 Bramhall Street, Maine Medical Center, Portland, ME 04102.

pain has on the patient's functioning. Among the factors that affect illness behavior are cultural background, socioeconomic status, psychological functioning, experiences, memory, and learning.

Although this term has been more frequently applied to chronic pain, extensive evidence indicates that the same factors also play a significant role in acute pain. In a pioneering study, Beecher observed that soldiers severely wounded in battle often did not complain of pain, or else described pain far below that which would be expected. He postulated that this may have reflected the relief felt by injured soldiers, who realized that they would be removed from the battlefield and that their lives would no longer be endangered (10). Conversely, stressful life events may contribute to the development of chronic pain (11).

A person's knowledge of and expectations regarding a potentially painful insult to the body have also been demonstrated to alter patient response. Egbert found that preoperative psychological preparation of the patient can have a profound influence on postsurgical recovery, including the degree of pain experienced by the patient (12).

Other work has indicated that various cultures and ethnic groups differ significantly with regard to level of pain reported and response to pain (13). What is considered to be acceptable illness behavior varies from society to society.

Patients' previous interactions with caregivers also have proven important in maintaining pain and in how patients will respond to future treatment efforts. Chronic pain frequently results in patients going from physician to physician seeking a "cure" for their problems and becoming dissatisfied with the health care system when no such solution is offered. This, in turn, makes it increasingly difficult to treat these patients (14).

Over the years, a variety of theories have been promulgated to explain pain. Although no single concept is universally accepted, the gate control theory developed by Melzack and Wall has received much attention (15). They believed that transmission of nerve impulses from the periphery to the spinal cord is modified by a gatelike mechanism in the dorsal horn. The position of the gate and the amount of information subsequently conveyed to the brain are determined by several factors. Large A-beta fibers as well as small A-delta and C fibers carry impulses from the periphery to the substantia gelatinosa and spinal cord transmission (T) cells. When the large fibers are activated, transmission to the T cells is inhibited, thus closing the gate; activation of the small fibers increases transmissions (opens the gate). The impact of the large and small fibers on the T cells is mediated by the substantia gelatinosa.

In addition to the impulses from the periphery, the gating mechanism is also influenced by descending messages from the brain and by a "central control" mechanism, which is activated by the L fibers and involves certain cognitive processes. Thus, according to this theory, pain is determined not only by peripheral stimulation, but also by information traveling from the brain to the spinal cord, reaffirming the role of the mind-body interaction.

Recent research indicates that the gating system is more complicated than first conceived and that more than one such mechanism may be involved, but the basic concept remains intact (16).

**Psychiatric Disorders and Pain.** Pain has been found to be associated with virtually all forms of psychiatric illness and to be a common initial symptom among many patients who have mental disorders (17, 18). Despite this, the American Psychiatric Association's *Diagnostic and Statistical Manual,* third edition, revised (DSM-III-R), the diagnostic classification of psychiatric disorders most frequently used in the United States, contains only one diagnosis that specifically addresses the problem of pain: somatoform pain disorder (19). The diagnostic criteria for this disorder are "preoccupation with pain for at least six months and either (1) appropriate evaluation uncovers no organic pathology or pathophysiologic mechanism (e.g., a physical disorder or the effects of injury) to account for the pain or (2) when there is related organic pathology, the complaint of pain or resulting social or occupational impairment is grossly in excess of what would be expected from the physical findings" (19).

This diagnosis and its DSM-III predecessor, psychogenic pain disorder (20), have been found to have limited usefulness because they exclude many patients who have both physical pathology and psychological problems related to their pain. Furthermore, it forces physicians to determine the level of pain that appropriately correlates with a particular degree of organic pathology—a difficult, if not impossible, task because the relationship between pain and physical impairment is poorly understood. Additionally, neither the presence nor the severity of psychiatric symptoms indicates whether organic pathology is present (21, 22).

Of all the psychiatric disorders, depression, in its various forms, has been identified as the most common problem among chronic pain patients (23, 24). It has even been speculated that chronic pain may actually be a variant of depression (25). Evidence to support a possible association includes the efficacy of the tricyclic antidepressant medications in the treatment of chronic pain, even in the absence of identifiable depression, and the increased frequency of chronic pain among patients who have a family history of depression (26, 27).

One of the difficulties in the evaluation of coexisting chronic pain and depression is that both entities can contribute to mutual symptoms, including change in appetite, impairment of sleep, anhedonia, and fatigue. The dexamethasone suppression test, a neurochemical marker for depression, can be useful in differentiating between the two. Patients with both depression and chronic pain are more likely to respond abnormally to the dexamethasone suppression test than are those with chronic pain alone (28). However, the usefulness of this test is limited by the numerous physical conditions and medications that can invalidate its results.

**Pain Measurement.** Because pain is a complex, subjective experience, many different methods for measuring it have been promulgated (29). Although no single method has been found to be universally valid or reliable, the following are among the most commonly used today.

The simplest form of pain measurement is the Numerical Rating Scale, for which the patient is asked to assign a numerical score to the pain. A typical scale ranges from 0 to 10, where 0 represents no pain and 10 corresponds to the worst pain imaginable. The Visual Analog Scale requires the patient to mark a place on a 10-cm line, the ends of which are similarly labeled (30).

The McGill Pain Questionnaire offers a more in-depth analysis of the pain the patient is experiencing (31). The test lists 20 sets of words that describe pain. These words are assigned to sensory, affective, and evaluative scales. The test can be scored based on the total number of words chosen or by the rank order of the words. Research indicates that patterns of response vary according to the type of pain experienced (32). The McGill Pain Questionnaire has been criticized for its reliance on language skills. Consequently, results may reflect the patient's intelligence level, education, or cultural background.

A commonly employed alternative is the West Haven–Yale Multidimensional Pain Inventory, which measures the patient's perception of how others respond to his or her pain; participation in daily activities; and overall life style (33).

Other, more traditional psychological tests have also been employed with pain patients. Among the most commonly used is the Minnesota Multiphasic Personality Inventory. However, the validity of this and similar psychological instruments, which were standardized among nonmedically-ill people, is controversial when applied to patients with physical ailments. For example, it was reported that patients whose back pain did not have an organic cause were more likely to demonstrate a certain configuration on this test, the so-called conversion V or neurotic triad, in which evaluations on the hypochondriasis and hysteria scales were believed to reflect the patient's concerns about his or her health. A lower depression score suggested that the patient was indifferent to these concerns (34). However, other researchers believe that this configuration may actually reflect adjustment to chronic illness and can be found among patients with chronic health problems, whether or not those problems have an identifiable organic cause (35, 36).

## Methods of Treatment

Psychologically based treatment approaches are considered to be a vital part of pain management programs (37). The most commonly used approaches fall into one of five categories: operant conditioning, biofeedback, cognitive-behavioral, relaxation training, and hypnosis.

**Operant Conditioning.** Operant conditioning is based on the concept that certain operant or learned behaviors develop in response to environmental cues. Common examples among pain patients include complaints of pain and reluctance to indulge in certain activities. The anticipated response to pain is often receipt of medication, dismissal from work, or relief from normal daily tasks. The goal of operant conditioning is to reinforce the positive or "healthy" behaviors and to diminish the destructive behaviors that maintain the patient's pain.

**Biofeedback.** In biofeedback, electronic equipment is employed to measure certain physiologic functions of which the patient is usually unaware, and to convey this information back to the patient. One example of such physiologic parameters is muscle tension. The patient may be instructed in specific relaxation techniques or encouraged to achieve certain goals involving instrumentation, such as lowering the pitch of a tone.

**Cognitive-Behavioral Therapy.** Cognitive-behavioral therapy involves identifying and correcting the patient's distorted attitudes, beliefs, and expectations. The goal of this therapy is to make the patient more aware of factors that exacerbate and diminish pain and to modify behavior accordingly.

A wide variety of relaxation techniques can be taught to patients. Among the most common methods is progressive muscle relaxation; the patient learns to relax different muscle groups by contracting and then relaxing each one (38).

**Hypnosis.** Athough the exact nature of hypnosis and the associated trance remain the subject of debate, a British Medical Association report offered the following apt description:

> A temporary condition of altered attention in the subject which may be induced by another person and in which a variety of phenomena may appear spontaneously or in response to verbal or other stimuli. These phenomena include alterations in consciousness and memory, increased susceptibility to suggestion, and the production in the subject of responses and ideas unfamiliar to him in his usual state of mind. Further, phenomena such as anesthesia, paralysis and rigidity of muscles, and vasomotor changes can be produced and removed in the hypnotic state. (39)

Although hypnosis involves relaxation, its efficacy in the treatment of pain extends beyond this. Patients can be taught to reduce the pain through hypnotic suggestions, such as forming a visual image of the pain and changing it, or dissociating the painful part from the rest of the body. Unfortunately, although much anecdotal evidence indicates the benefits of hypnosis in the treatment of pain, there are relatively few controlled studies (40).

**Applying These Approaches.** The application of each of these techniques to the management of pain has been studied to varying degrees, and controversy still remains as to which approach is most efficacious. However, the research indicates certain general guidelines on the best use of each technique.

In extensive literature reviews, Turner and Chapman (41, 42) and Linton (43) found operant therapy to be especially useful in decreasing patients' medication use and increasing their activity levels. In contrast, cognitive-behavioral therapy was observed to be helpful in reducing pain complaints.

Each of these reviews found biofeedback to be of benefit in certain pain conditions, especially tension and migraine headaches, but, overall, to be inferior or at best equal to relaxation therapy, which appears to be quite efficacious in the treatment of a wide variety of both acute and chronic pain conditions. Furthermore, biofeedback may be detrimental in some cases of chronic pain by focusing only on the somatic aspects of pain. Additionally, the practitioner must consider the cost of biofeedback instrumentation, which can range from a hundred to thousands of dollars.

Each of these therapies is applicable where pain is the primary problem. However, when pain is the symptom of a mental disorder, appropriate treatment with psychotherapy or psychotropic medications must be initiated.

One additional method of treatment, the efficacy and actions of which are controversial, is the use of placebo. Research indicates that placebos can have a significant analgesic effect in a variety of painful conditions, including those having an identifiable organic etiology (44).

## Summary

Psychological factors have a major role in the development and maintenance of pain. Further research to clarify the nature of the relationship between these factors and pain is required. When evaluating patients with acute or chronic pain, attention must be given to psychological factors.

### References

1. Ostrerweis M, Kleinman A, Mechanic D, eds. Pain and disability. Washington, DC: National Academy Press, 1987.
2. Sapira JD. In: Moskow SB, ed. Hunan hand and other ailments. Boston: Little Brown, 1987, 68.
3. Chapman CR, Turner JA. Psychological control of acute pain in medical settings. J Pain Symptom Management 1986; 1:9–20.
4. King SA, Strain JJ. Benzodiazepines and chronic pain. Pain 1990; 41:3–4.
5. Black RG. The chronic pain syndrome. Surg Clin N Am 1975; 55:999–1011.
6. Tunks E. Is there a chronic pain syndrome? In: Lipton S, Tunks E, Zoppi M eds. Advances in pain research and therapy, vol. 13. New York: Raven, 1990, 257–266.
7. Bonica JJ. Definitions and taxonomy of pain. In: Bonica JJ, ed. The management of pain, 2nd ed. Philadelphia: Lea & Febiger, 1990.
8. Mechanic D. The concept of illness behavior. J Chron Dis 1962; 15:189–194.
9. Pilowsky I. Abnormal illness behavior. Br J Med Psychol 1968; 42:347–351.
10. Beecher HK. Pain in men wounded in battle. Ann Surg 1946; 123:98–105.
11. Atkinson JH, Slater MA, Grant I, Patterson TL, Garfin SR. Depressed mood in chronic low back pain: relationship with stressful life events. Pain 1988; 35:47–55.
12. Egbert LD, Battit GE, Welch CE, Bartlett MK. Reduction of postoperative pain by encouragement and instruction of patients. N Engl J Med 1964; 270:825–827.
13. Zborowski M. People in pain. San Francisco: Jossey-Bass, 1969.
14. Moulton RE. Emotional factors in non-organic temporomandibular joint pain. Dent Clin North Am 1966; 609–620.
15. Melzack R, Wall PD. Pain mechanisms: a new theory. Science 1965; 150:971–979.

16. Melzack R, Wall PD. The challenge of pain. New York: Basic Books, 1982.
17. Dworkin RH, Caligor E. Psychiatric diagnosis and chronic pain: DSM-III-R and beyond. J Pain Symptom Management 1988; 3:87–98.
18. Pilling LF, Brannick TL, Swenson WM. Psychological characteristics of psychological patients having pain as a presenting problem. Can Med Assoc J 1967; 97:387–394.
19. American Psychiatric Association. Diagnostic and statistical manual of mental disorders. 3rd ed, revised. Washington, DC: American Psychiatric Association, 1987:266.
20. American Psychiatric Association. Diagnostic and statistical manual of mental disorders. 3rd ed. Washington, DC: American Psychiatric Association, 1980.
21. Benjamin S, Barnes D, Berger S, Clarke I, Jeacock J. The relationship of chronic pain, mental illness and organic disorders. Pain 1988; 32:185–195.
22. Trief PM, Elliott DJ, Stein N, Frederickson BE. Functional vs. organic pain: a meaningful distinction? J Clin Psychol 1987; 43:219–226.
23. King SA, Strain JJ. The problem of psychiatric diagnosis for the pain patient in the general hospital. Clin J Pain 1989; 5:329–335.
24. Romano JM, Turner JA. Chronic pain and depression: does the evidence support a relationship? Psychol Bull 1985; 97:18–34.
25. Blumer D, Heilbronn M. Chronic pain as a variant of depressive disease: the pain-prone disorder. J Nerv Ment Dis 1982; 170:381–394.
26. Feinmann C. Pain relief by antidepressants: possible modes of action. Pain 1985; 23:1–8.
27. Katon W, Egan K, Miller D. Chronic pain: lifetime diagnoses and family history. Am J Psychiatry 1985; 142:1156–1160.
28. France RD, Krishnan KRR, Trainor M, Pelton S. Chronic pain and depression. IV. DST as a discriminator between chronic pain and depression. Pain 1987; 28:39–44.
29. Chapman CR, Casey KL, Dubner R, Foley KM, Gracely RH, Reading AE. Pain measurement: an overview. Pain 1985; 22:1–31.
30. Scott J, Huskisson EC. Graphic representation of pain. Pain 1976; 2:175–184.
31. Melzack R. The McGill Pain Questionnaire: major properties and scoring methods. Pain 1975; 1:277–299.
32. Reading AE, Everitt BS, Sledmere CM. The McGill Pain Questionnaire: a replication of its construction. Br J Clin Psychol 1982; 21:339–349.
33. Kerns RD, Turk DC, Rudy TE. The West Haven–Yale Multidimensional Pain Inventory (WHYMPI). Pain 1985; 23:345–356.
34. Hanvik LJ. MMPI profiles in patients with low-back pain. J Consult Clin Psychol 1951; 15:350–353.
35. Naliboff BD, Cohen MJ, Yellen AN. Does the MMPI differentiate chronic illness from chronic pain? Pain 1982; 13:333–341.
36. Watson D. Neurotic tendencies among chronic pain patients: an MMPI item analysis. Pain 1982; 14:365–385.
37. Fordyce WE, Roberts AH, Sternback RA. The behavioral management of chronic pain: a response to critics. Pain 1985; 22:113–125.
38. Jacobson E. Modern treatment of tense patients. Springfield, IL: Charles C Thomas, 1970.
39. British Medical Association Report. Medical use of hypnotism. Br Med J 1955; I(Suppl):190–193.
40. Hilgard ER, Hilgard JR. Hypnosis in the relief of pain, revised edition. Los Altos, CA: William Kaufman, 1983.
41. Turner JA, Chapman CR. Psychological interventions for chronic pain: a critical review. I. Relaxation training and biofeedback. Pain 1982; 12:1–21.
42. Turner JA, Chapman CR. Psychological interventions for chronic pain: a critical review. II. Operant conditioning, hypnosis, and cognitive-behavioral therapy. Pain 1982; 12:23–46.
43. Linton SJ. Behavioral remediation of chronic pain: a status report. Pain 1986; 24:125–141.
44. Evans FJ. The placebo response in pain control. Psychopharmacol Bull 1981; 17:72–76.

# Nociceptive Transmission within the Spinal Cord

LINDA S. SORKIN, PH.D.

ACUTE pain is a subjective experience that normally results from activation of peripheral nociceptors. It is indicative of potential or actual tissue damage. Pain perception is frequently divided into two components: the sensory-discriminative aspect, which describes the quality and location of the pain (1), and the motivational-affective portion, which leads to aversion and avoidance. As the pain message ascends in the central nervous system, the pathways responsible for these two components become anatomically distinct (2).

**Primary Afferent Fibers.** Nociceptors are unencapsulated "free" nerve endings that fire when mechanical, thermal, or chemical stimuli threaten or produce damage. They are located in skin as well as in viscera and muscle, on blood vessels, and within joints. All nociceptors are not activated by all types of nociceptive stimuli. A subpopulation of fine myelinated (Aδ) and unmyelinated (C) nerve fibers innervate nociceptors and transmit information concerning damage to the spinal cord. Activation of the Aδ fibers produces a short-duration, well-localized pain; in contrast, activation of the C fibers produces a poorly localized, burning sensation (3, 4). Pain sensation is dependent on the connectivity of the fibers being activated, not on the pattern of stimulation, as demonstrated in humans by microneurography; recordings are made from single nerve fibers and the activating stimuli determined. When nociceptive fibers are stimulated electrically through the microelectrode, a sensation of pain (3, 5) is produced. When other fibers are stimulated at the same or higher frequencies or intensities, under normal conditions pain is never produced; the sensations are those appropriate to the receptor type (6, 7).

Many of the small fibers contain combinations of excitatory amino acids and neuropeptides (8–12). Noxious stimulation of the tissue or C-fiber intensity electrical stimulation of nerves leads to selective release of several putative neurotransmitters both centrally into the spinal cord (13–17) and peripherally into the tissue (18). It is thought that peripheral release of peptides from primary afferent fibers contributes to the development of inflammation (19, 20). Electrical stimulation of the sciatic nerve also decreases or depletes the amount of substance P (SP), calcitonin gene-related peptide (CGRP), and cholecystokinin-8 found within the nerve fibers and their terminals within the dorsal horn (21, 22).

Experiments in animal models show a correlation between the stimulus used to induce pain (that is, the type of pain receptors activated) and the particular agent released into the cord. Increased levels of substance P in the spinal cord are found after application of noxious heat or mechanical stimuli to the skin (15, 23). In contrast, only noxious thermal stimulation leads to somatostatin release (14). Snow and colleagues, however, found no consistent pattern between immunoreactive substances found within a nerve fiber and the modalities that activate it (9).

The fine primary afferent fibers enter the spinal cord in the lateral dorsal root and then bifurcate within Lissauer's tract to travel both rostrally and caudally. In addition, a significant numbr of unmyelinated fibers are found within the ventral root (24). This has been demonstrated in several species, including humans. The termination pattern of a fiber is a function of both its

---

This work was suported by NIH grant NS 11255 and the Bristol-Myers Squibb Company. From the Department of Anatomy and Neurosciences and Marine Biomedical Institute, University of Texas Medical Branch, Galveston, Texas. Address reprint requests to Linda S. Sorkin, Ph.D., Department of Anatomy and Neurosciences, Marine Biomedical Institute, University of Texas Medical Branch, Galveston, TX 77550.

modality and source. The heaviest concentration of terminals from physiologically identified "pain fibers" is located ipsilaterally in Rexed's laminae I, II, and V, contralaterally in lamina V, and around the central canal (25–27). The visceral afferents are widespread, extending for up to five spinal segments, compared to one or two segments for afferents innervating the skin (27). This opportunity for divergence throughout several spinal segments may contribute to the diffuse quality of visceral pain.

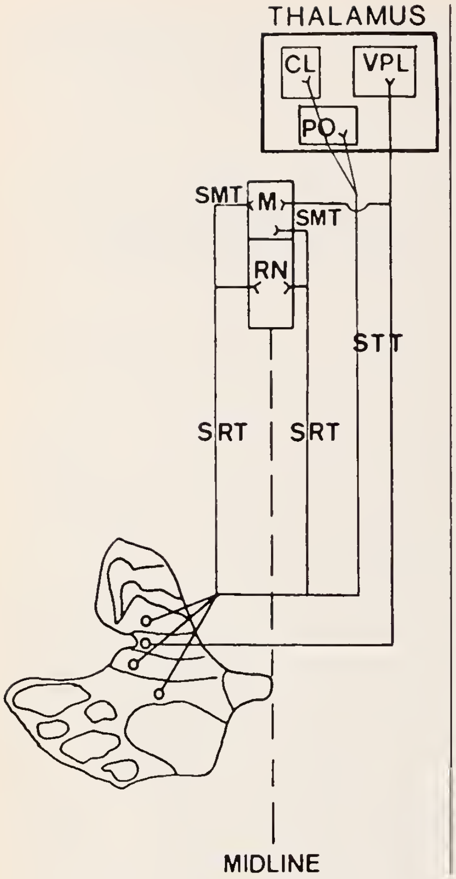
**Spinal Cord.** After the nociceptive signal enters the spinal cord, it is relayed to the brain via projection or tract cells; connections are both mono- and polysynaptic. Processing of the signal within the spinal cord can occur via primary afferent inhibition, actions on interneurons, or direct modulation of the tract cells. Local, intrasegmental, and descending modulatory sources are involved. In the absence of pharmacologic intervention, modulation is probably dependent on behavioral state as well as prior activation history of the cell.

Electrophysiologic examination of dorsal horn cells with cutaneous nociceptive inputs gives rise to two major categories: nociceptive-specific (NS) and wide-dynamic-range (WDR), or multireceptive, cells. NS cells respond only to noxious stimulation. WDR cells respond to both innocuous and noxious stimulation of their receptive fields, and their firing rate is correlated with stimulus intensity. Firing of NS cells is a clear message that nociceptors have been stimulated. Discharge of a WDR cell may only be interpreted in context. Concentrations of these cells are found in lamina I (predominantly NS) and lamina V (predominantly WDR), corresponding to the location of the finely myelinated and unmyelinated nerve terminals. There is evidence that the division between the two types of cells is labile and that under different modulatory conditions, cells may be classified in different categories (28).

**Spinal Pathways that Transmit the Nociceptive Message.** Postmortem examination of lesions produced by injury or disease indicates that in humans, somatic and visceral pain transmission is most strongly correlated with the integrity of the contralateral anterolateral quadrant of the spinal cord (29). In addition, electrical stimulation of the anterolateral quadrant in awake patients usually produces reports of contralateral pain. Anterolateral cordotomy produces immediate relief of noncentral pain in 70%–80% of patients (30–32). Sectioning the entire ventrolateral quadrant, up to a few millimeters from the anterior median fissure, is far more effective than superficial lesions involving only the anterolateral funiculus. Unfortunately, pain frequently returns after cordotomy, and the one-year success rate falls to approximately 50% (30, 31). Intense noxious peripheral stimulation can sometimes be perceived by the "analgesic" patient; thus, the analgesia might more correctly be termed "hypalgesia." These clinical observations have been replicated in the subhuman primate (33). After unilateral cordotomy, the animal's response to pain always recovers. Even after bilateral section of both the ventral and ventrolateral columns, after several months, responses to noxious stimulation return. Percutaneous cordotomies restricted to the dorsolateral funiculus have produced mixed results, in some cases causing complete analgesia (34) and in others having no apparent effect on pain sensibility (35, 36).

Experimentally, cells from several neuronal tracts, located in both the ventral and dorsal funiculi, have been shown to respond to noxious stimulation. One explanation of "recovery" after cordotomy is that these alternate intact pathways gradually take over the function of those that have been sectioned. When anterolateral cordotomy is successful, the analgesic area is almost always contralateral to the lesion (37) and one to two segments caudal to it. The later characteristic is more variable. Commissural myelotomy produces a bilateral analgesic band around the body, and the region of clinical pain loss may extend further rostral or caudal than would be predicted by the anatomy of the ascending systems (38, 39).

**Ventral Pathways.** There are three primary pain transmission pathways in the anterolateral spinal cord; all have major crossed components. They are the spinothalamic (STT), spinoreticular (SRT), and spinomesencephalic (SMT) tracts (Fig. 1). Many of the dorsal horn projection cells send collaterals to more than one pathway. Cells contributing to the STT are located throughout the dorsal and ventral horns of the spinal cord (40). Axons from approximately 90% of them cross the spinal cord, and the crossing is within one segment of the cell body. Axons from STT cells are arranged somatotopically in the ventrolateral funiculus and ascend to the contralateral thalamus (41, 42). Kuru (43) divided these into the ventromedial and dorsolateral spinothalamic tracts. Results from cordotomies indicate that in most cases, both these tracts must be lesioned to produce adequate pain relief. Medially projecting cells terminate in the nucleus submedius of the thalamus as well as the central lateral and other medial nuclei. The ventromedial tract cells of or-



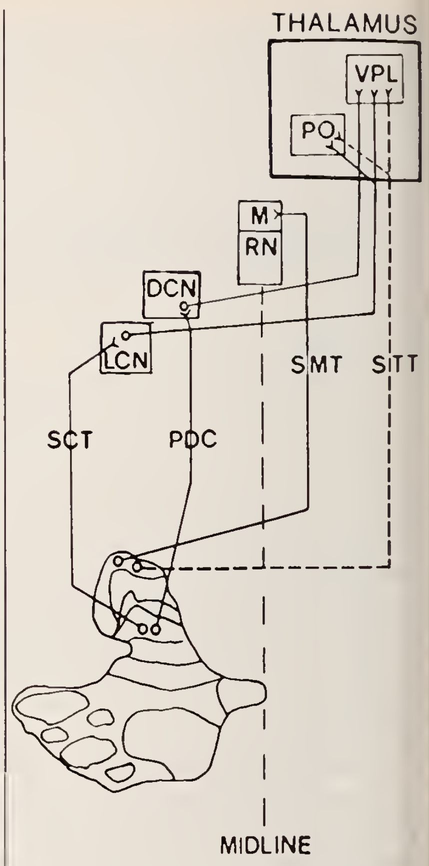


Fig. 1. **Left** Pain pathways in ventrolateral quadrant of spinal cord. Spinothalamic tract (STT) terminates in central (CL) and posterior complex (PO) of medial thalamus as well as ventral posterior lateral (VPL) nucleus of lateral thalamus. Spinomesencephalic (SMT) and spinoreticular (SRT) tracts project throughout midbrain (M) and caudal reticular nuclei (RN). SMT and SRT pathways are predominantly contralateral with a small but significant ipsilateral component. Fig. 2. **Right** Pathways carrying nociceptive information in mid- and dorsal spinal cord. Lamina I STT cells terminate in PO and VPL thalamic nuclei. Other cells in lamina I project to midbrain. Spinocervical tract (SCT) cells project through dorsolateral fasciculus to lateral cervical nucleus (LCN). This nucleus lies ventrolateral to dorsal horn of $C_1$–$C_3$ spinal segments. A fourth dorsal tract projects through dorsal columns to dorsal column nuclei (DCN). PDC, postsynaptic dorsal column.

igin are located primarily in laminae IV–VIII (44). Cells with axons projecting laterally to the ventral posterior lateral nucleus of the thalamus (VPL), the major sensory relay nucleus of the thalamus, are concentrated in lamina V. Distributions of the these two cell populations are overlapping and some cells send collaterals to both medial and lateral thalamus. The intralaminar nuclei project to limbic regions and frontal and parietal cortex. The lateral division, parallel to other somatosensory pathways, projects to areas SI and SII of the sensory cortex (45). This medial-lateral split clearly represents the division between the affective and discriminative components of pain.

Spinothalamic tract cells that project exclusively to the medial thalamic nuclei are predominantly NS, with large, frequently whole-body receptive fields. Many of these cells continue to discharge after the cessation of noxious stimulation (44). The majority of STT cells projecting to VPL respond to high-intensity stimulation and presumably are involved in nociception. Most are WDR (60%) and NS (32%), the remainder responding exclusively to low-intensity or proprioceptive stimuli (46). Receptive fields are ipsilateral to the cell body, are usually excitatory, and are relatively small. Receptive fields of NS cells are significantly smaller than those of WDR cells (47). Evoked responses in many of these cells are inhibited by noxious stimulation from many other parts of the body. Supraspinal pathways are involved because high spinal transection reduces the effect of this inhibitory control system (48). Thus, cells with well-localized receptive fields, which also code for intensity within the noxious range, feed into the laterally located sensory-discriminative system. Cells with little spatial and temporal specificity, which signal the occurrence of damage, project medially into a system associated with arousal and avoidance.

Throughout the length of the spinal cord, STT cells are found that respond to stimulation of muscle and viscera in addition to cutaneous structures (49, 50). Many STT cells in the upper thoracic cord respond to stimulation of cardiopulmonary visceral afferent fibers (51–53) as well as to occlusion of a coronary artery. Responses to occlusion may be delayed, indicating that the actual stimulus is the development of ischemia in cardiac muscle (54). Other STT cells in thoracic cord can be activated by distention of the gall bladder (55), kidney (56), or ureter (57). STT cells in the upper lumbar and sacral segments receive input from the bladder and testes (50). VPL projecting cells with input from each organ have cutaneous receptive fields within a particular distribution. The location of these fields corresponds to the site of referred pain in people with visceral disease. This cutaneous-visceral convergence provides an explanation for pain referral.

In the primate, spinoreticular tract SRT cells project bilaterally (58). Their response properties and receptive field characteristics may resemble either type of STT cell (medially or laterally projecting), but their cells of origin are in the same region of the spinal cord as the medially projecting STT cells (59). They terminate in the more medial of the reticular nuclei. Most SRT cells are maximally activated by nociceptive stimuli. The majority of SRT cells with visceral input are inhibited by vagal stimulation. Spinomesencephalic tract (SMT) cells also project bilaterally (60, 61); the ipsilateral projection is a minor one. These cells frequently have large excitatory receptive fields with complex inhibitory input from skin, muscle, and viscera (62, 63). The cells of origin are located in the same part of the cord as the VPL projecting cells (61). When cells have projections to both VPL thalamus and the mesencephalon, they have characteristics resembling those of lateral thalamic projecting cells, for example, excitatory receptive fields confined to a single limb (63).

**Dorsal Pathways.** Strictly speaking, the pathways called "dorsal" might more correctly be referred to as pain pathways not in the anterolateral funiculus. The cells of origin of these pathways have response properties similar to those of ventrally projecting cells (NS and WDR cells). These pathways probably contribute to the pain experience in the intact organism and may take on some of the functions of the ventral pathways after injury. These alternate dorsal pathways include the STT and SMT projections from lamina I, the spinocervicothalamic (SCT) pathway, and the postsynaptic dorsal column pathway (PDC) (Fig. 2).

Lamina I contains both NS (the majority of the cells) and WDR cells that project to thalamus and the midbrain (64–66). Animal experiments originally indicated that these pathways traveled in the dorsolateral fasciculus (DLF) rather than in the ventrolateral quadrant (67); more recent work by Craig (68) indicates that the pathways are in the middle of the lateral funiculus. Either way, they are distinctly dorsal to the projections from the mid- and deep dorsal horn. Spinocervical tract (SCT) cells are concentrated in lamina IV, respond to noxious stimulation, and project to the ipsilateral lateral cervical nucleus (LCN) through the DLF (69). The pathway then projects to the contralateral thalamus (70, 71). In monkeys, slightly more than half the LCN cells are either WDR or NS cells (71). In contrast, the ventral and lamina I STT projections contain 92% and 100% pain cells, respectively (46). Cells in the postsynaptic dorsal column pathway are located in lamina IV and project within the deep dorsal columns to the ipsilateral dorsal column nuclei (72–74); many are WDR cells. Although the medial lemniscus projects from the DCN to the contralateral thalamus, it is uncertain to what extent nociceptive information is transmitted via this pathway. Not much work has been done on this system in the primate (75).

**Modulation of Pain within the Spinal Cord.** Pain can be pharmacologically blocked within the spinal cord by inhibiting the release of neurotransmitters from the small-diameter primary afferent fibers, administering antagonists to primary afferent neurotransmitters, or inhibiting transmission of the pain message either at the tract (for example, STT or SMT) cells or within interneuronal processing circuits.

Presynaptic inhibition of neurotransmitter release from primary afferents requires that synapses from descending or intrinsic elements of the endogenous analgesia system terminate on the primary afferent nerve terminals. This criterion has been met by the $\mu$ and $\delta$ opiate receptors (76), serotonin ($5HT_3$) (77), and $\gamma$-aminobutyric acid (GABA) systems (78). Administration of opiates, noradrenergic agonists, and drugs that increase spinal levels of endogenous monoamines decreases the release of SP and/or CGRP (79–83). Thus, this represents the first site at which the pain message is modified.

Antagonists to only a few primary afferent neurotransmitters have been given intrathecally to awake-behaving animals. D-amino acid–substituted forms of SP alter the nociceptive threshold of rats (84–86). However, these antagonists are also partial agonists. In addition, the doses required for analgesia are just slightly less than those that produce temporary flaccid paralysis (84). Increases in the dose lead to permanent loss of motor function, and histologic examination of the spinal cord shows extensive necrosis in both the dorsal and ventral horns (85). Measurements of blood flow after administration of SP antagonists suggest that necrosis may be secondary to spinal ischemia (87). Intrathecal administration of antagonists to a subtype of excitatory amino acid receptor (NMDA) produces analgesia when injected by itself and reverses the hyperalgesia produced by intrathecal amino acid agonists (88). $\text{GABA}_A$ agonists also selectively block hyperalgesia induced by excitatory amino acids (89), and $\text{GABA}_B$ agonists block the behavioral responses associated with intrathecal SP (90).

Animal studies have shown that local application of several putative neurotransmitters (opiates, 5HT, norepinephrine [NE], GABA, glycine, and acetylcholine) inhibits the responses of dorsal horn cells (91–93). Some of this modulation could be presynaptic; much of it is known to be postsynaptic. However, administration of many of these neurotransmitters causes excitation or an enhancement of responsiveness in some dorsal horn cells. This discrepancy may be due to cells having different projections or functions (projection cell or inhibitory interneuron) or differences in the part of the cell (soma or far dendritic tree) accessed by the drug (94, 95).

The spinal actions and interactions of 5HT and norepinephrine have been studied in depth. Brainstem nuclei are the sources of both monoamines; the descending fibers travel in the lateral funiculus (96–98). Noxious peripheral stimulation leads to release of both neurotransmitters within the spinal cord (99–101). Chemical lesions of either system lower the animal's threshold to painful input (102, 103). This is consistent with the idea that there are tonically active serotonergic and noradrenergic pathways that modulate pain processing at the level of the spinal cord.

All of the 5HT within the spinal cord is thought to originate within the caudal raphe nuclei. Approximately 80% of the raphe spinal fibers contain 5HT; many contain colocalized neurotransmitters/neuromodulators, including GABA, substance P, enkephalin, and other peptides (104, 105). Direct stimulation of the raphe nuclei evokes the release of 5HT in the spinal cord and raises the threshold of nocifensive (aimed at avoiding noxious stimuli) reflexes (106, 107). Analgesia produced by intrathecal 5HT occurs without motor side effects, although greatly increased levels of 5HT cause tremor and a mild monoclonus (103).

Lesions of the raphe nuclei lower the threshold for nocifensive reflexes (108). Intrathecal administration of 5HT antagonists also causes hyperalgesia and prevents or reduces analgesia associated with raphe stimulation and intrathecal 5HT (103). Depleting endogenous stores of 5HT by feeding rats a tryptophan-free diet results in hypersensitivity to stimuli (109).

There is still controversy regarding interactions of the serotonergic system with morphine administered systemically. Depletion of 5HT and raphe nuclei lesions reduces the effectiveness of morphine (109, 110). Conversely, intrathecal administration of naloxone reduces the effectiveness of intrathecal 5HT (111). Increasing the availability of 5HT is thought to potentiate the actions of morphine, although this opinion is not unanimous (112).

The serotonergic analgesia system is dependent on the integrity of the noradrenergic pathways. Systemic depletion of NE or blockade of α-adrenergic terminals interferes with or prevents the hypoalgesia normally caused by intrathecal 5HT or electrical stimulation within the caudal raphe nuclei (113–115). However, depletion of 5HT appears to have no effect on adrenergic-related analgesia (116).

Catecholamine cells projecting to the spinal cord originate primarily in the locus coeruleus-subcoeruleus complex, nucleus Kolliker-Fuse, lateral reticular nucleus, and magnocellular region of the reticular formation (117). Stimulation of these sites leads to increased levels of NE and NE metabolites within the spinal cord (107, 118, 119), inhibits the responses of dorsal horn neurons to high-intensity stimulation, and inhibits nocifensive reflexes (119, 120). Intrathecal NE elevates pain thresholds (121, 122). Analgesia produced by supraspinal stimulation and intrathecal NE appears to be mediated by the α-adrenergic receptor, principally the α-2 receptor (120). Intrathecally administered noradrenergic antagonists cause hyperalgesia (123). The α-2 agents are more effective than the α-1, but α-1 agonists and not α-2 reverse NMDA-induced nociceptive activity (88).

Lesions within the noradrenergic system also interfere with morphine analgesia (124). NE-depleting agents, including reserpine, as well as α-2 antagonists, block morphine analgesia (123, 125). However, unlike the serotonergic system, pretreatment with naloxone does not block noradrenergic analgesia (121). The analgesia is not mimicked by intrathecal vasoconstrictors, nor is it antagonized by local administration of vasodilator agents (103, 121). This indicates that the changes in responsiveness are not the result of local spinal ischemia.

Intrathecal clonidine, an α-2 agonist, raises nociceptive thresholds in animals. Clinically, epidural and intrathecal clonidine are used to provide analgesia (126). When these agents are given in conjunction with opiates, there appears to be a synergistic action that extends both the magnitude and duration of the analgesia. Combination therapy also delays the development of tol erance to either drug. Clonidine can prevent many of the symptoms of morphine withdrawal (127, 128).

The analgesic effects of 5HT, norepinephrine, and their agonists are potentiated by prior administration of monoamine oxidase (MAO) inhibitors and monoamine reuptake inhibitors. These drugs lead to an increase in the extracellular concentrations of 5HT and NE in the spinal cord and increased pain thresholds (103). Tricyclic antidepressants, in addition to their antidepressant effects, relieve pain; this may be a function of their ability to block reuptake of monoamines (129). Selective 5HT reuptake inhibitors are also being used to treat chronic pain (130). Some local anesthetics (esters) are MAO inhibitors, another class (amides) are not. Tetracaine, an ester, also inhibits reuptake of NE into synaptosomes. Thus ester local anesthetics increase available monoamine by both mechanisms (131). Both MAO inhibitors and reuptake inhibitors also potentiate the analgesic actions of morphine (129, 132). This potentiation is blocked by administration of either 5HT or α-adrenergic antagonists.

**Summary.** Pain pathways in the spinal cord transmit information about actual or impending tissue damage. There are parallel pathways to both the medial (motivational-affective) and lateral (sensory-discriminative) portions of the system. Endogenous systems modulate pain transmission and ultimately pain perception. These modulatory systems can be accessed pharmacologically by naturally occurring neurotransmitters and neuromodulators. Research is being done to identify receptor specific agents that selectively inhibit pain and do not interfere with processing of other information.

## Acknowledgments

I wish to thank Dr. W. D. Willis for reading a copy of the manuscript.

## References

1. Melzack R, Casey KL. Sensory, motivational and central control determinants of pain. In: Kenshalo DR, ed. The skin senses. Springfield, IL: Charles C Thomas, 1968:423–443.
2. Price DD, Dubner R. Neurons that subserve the sensory discriminative aspects of pain. Pain 1977; 3:307–338.
3. Torbejork HE, Ochoa JL. Specific sensations evoked by activity in single identified units in man. Acta Physiol Scand 1980; 110:445–447.
4. Torbejork HE, Ochoa JL. Selective stimulation of sensory units in man. Adv Pain Res Ther 1983;5:99–104.
5. Konietzny F, Perl ER, Trevino D, Light A, Hensel H. Sensory experiences in man evoked by intraneural electrical stimulation of intact cutaneous afferent fibers. Exp Brain Res 1981; 42:219–222.
6. Scharf B, Hyvarinen J, Poranen A, Merzenich MM. Electrical stimulation of human hair follicles via microelectrodes. Percept Psychophysiol 1973; 14:273.
7. Vallbo AB. Sensations evoked from the glabrous skin of the human hand by electrical stimulation of unitary mechanosensitive afferents. Brain Res 1981; 215:359–363.
8. Skofitsch G, Jacobowitz DM. Calcitonin gene-related peptide coexists with substance-P in capsaicin sensitive neurons and sensory ganglia of the rat. Peptides 1985; 6:747–754.
9. Leah JD, Cameron AA, Snow PJ. Neuropeptides in physiologically identified mammalian sensory neurons. Neurosci Lett 1985; 56:257–264.
10. De Biasi S, Rustioni A, Glutamate and substance P coexist in primary afferent terminals in the superficial laminae of spinal cord. Proc Natl Acad Sci USA 1988; 85:7820–7824.
11. Westlund KN, McNeill DL, Patterson JT, Coggeshall RE. Aspartate immunoreactive axons in normal rat $L_4$ dorsal roots. Brain Res 1989; 489:347–351.
12. Westlund KN, McNeill DL, Coggeshall RE. Glutamate immunoreactivity in rat dorsal root axons. Neurosci Lett 1989; 96:13–17.
13. Go VLW, Yaksh TL. Release of substance P from the cat spinal cord. J. Physiol 1987; 391:141–167.
14. Kuraishi Y, Hirota N, Sato Y, Hino Y, Satoh M, Takagi H. Evidence that substance P and somatostatin transmit separate information related to pain in the spinal dorsal horn. Brain Res 1985; 325:294–298.
15. Brodin E, Linderoth B, Gazelius B, Ungerstedt U. In vivo release of substance P in cat dorsal horn studied with microdialysis. Neurosci Lett 1987; 76:357–362.
16. Skilling SR, Smullen DH, Beitz AJ, Larson AA. Extracellular amino acid concentrations in the dorsal spinal cord of freely moving rats following veratridine and nociceptive stimulation. J Neurochem 1988; 51:127–132.
17. Liu D, Sorkin LS, Hughes MG, McAdoo DJ. Release of aspartate and glutamate in the dorsal horn of the cat lumbar spinal cord in response to intradermal capsaicin injection. Neurosci Abstr 1989; 15:549.
18. Yaksh TL. Substance P release from knee joint afferent terminals: modulation by opioids. Brain Res 1988; 458:319–324.
19. Levine JD, Clark R, Devor M, Helms C, Moskowitz MA, Basbaum AI. Intraneuronal substance P contributes to the severity of experimental arthritis. Science 1984; 226:547–549.
20. Fitzgerald M. Arthritis and the nervous system. Trends Neurosci 1989; 12:86–87.
21. Klein CM, Coggeshall RE, Carlton SM, Westlund KN, Sorkin LS. Changes in calcitonin gene-related peptide immunoreactivity in the rat dorsal horn following electrical stimulation of the sciatic nerve. Neurosci Lett 1990; 115:149–154.
22. Klein CM, Sorkin LS, Carlton SM, Westlund KN, Coggeshall RE. Changes in peptide immunoreactivity in rat dorsal horn following electrical stimulation of the sciatic nerve. Neurosci Abstr 1989; 15:299.
23. Duggan AW, Morton CR, Zhao ZQ, Hendry IA. Noxious heating of the skin releases immunoreactive substance P in the substantia gelatinosa of the cat: a study with antibody microprobes. Brain Res 1987; 403:345–349.
24. Coggeshall RE, Ito H. Sensory fibers in ventral roots L7 and S1 in the cat. J Physiol 1977; 267:215–235.

25. Light AR, Perl ER. Spinal termination of functionally identified primary afferent neurons with slowly conducting fibers. J Comp Neurol 1979; 186:133–150.
26. Sugiura Y, Lee CL, Perl ER. Central projections of identified, unmyelinated (C) afferent fibers innervating mammalian skin. Science 1986; 234:358–361.
27. Sugiura Y, Terui N, Hosoya Y, Kohno K. Distribution of unmyelinated primary afferent fibers in the dorsal horn. In: Cervero F, Bennett GJ, Headley PM, eds. Processing of sensory information in the superficial dorsal horn of the spinal cord. New York: Plenum, 1989:15–27.
28. Simone D, Sorkin LS, Oh U, et al. Neurogenic hyperalgesia: central correlates in responses of spinothalamic track neurons. J Neurophysiol (in press).
29. Noordenbos W, Wall PD. Diverse sensory functions with an almost totally divided spinal cord. A case of spinal cord transection with preservation of part of one anterolateral quadrant. Pain 1976; 2:185–195.
30. Nathan PW. Results of antero-lateral cordotomy for pain in cancer. J Neurol Neurosurg Psychiatry 1963; 26:353–362.
31. White CW, Richardson EP, Sweet WH. Upper thoracic cordotomy for relief of pain: postmortem correlation of spinal incision with analgesic levels in 18 cases. Ann Surg 1956; 144:407–420.
32. White JC, Sweet WH, Hawkins R, Nilges RG. Anterolateral cordotomy: results, complications and causes of failure. Brain 1950; 73:346–367.
33. Vierck CJ, Luck MM. Loss and recovery of reactivity of noxious stimuli in monkeys with primary spinothalamic cordotomies, followed by secondary and tertiary lesions of other cord sectors. Brain 1979; 102:233–248.
34. Moffie D. Spinothalamic fibres, pain conduction and cordotomy. Clin Neurol Neurosurg 1975; 78:261–268.
35. Nathan PW, Smith MC, Cook AW. Sensory effects in man of lesions of the posterior columns and of some other afferent pathways. Brain 1986; 109:1003–1041.
36. Nathan PW. Comments on "A dorsolateral spinothalamic tract in macaque monkey" by Apkarian and Hodge. Pain 1990; 40:239–240.
37. Voris HC. Ipsilateral sensory loss following cordotomy. Report of a case. Arch Neurol Psychiatry 1951; 65:95–96.
38. Cook AW, Nathan PW, Smith MC. Sensory consequences of commissural myelotomy. A challenge to traditional anatomical concepts. Brain 1984; 107:547–568.
39. Hitchcock E. Stereotaxic cervical myelotomy. J Neurol Neurosurg Psychiatry 1970; 33:224–230.
40. Willis WD, Kenshalo DR Jr, Leonard RB. The cells of origin of the primate spinothalamic tract. J Comp Neurol 1979; 188:543–574.
41. Boivie J. An anatomical reinvestigation of the termination of the spinothalamic tract in the monkey. J Comp Neurol 1979; 186:343–370.
42. Applebaum EA, Beall JE, Foreman RD, Willis WD. Organization and receptive fields of primate spinothalamic tract neurons. J Neurophysiol 1975; 38:572–586.
43. Kuru M. Sensory paths in the spinal cord and brain stem of man. Tokyo: Sogensya, 1949.
44. Giesler GJ, Yezierski RP, Gerhart KD, Willis WD. Spinothalamic tract neurons that project to medial and/or lateral thalamic nuclei: evidence for a physiologically novel population of spinal cord neurons. J Neurophysiol 1981; 46:1285–1308.
45. Willis WD, Coggeshall RE. Sensory mechanisms of the spinal cord. New York: Plenum Press, 1978.
46. Willis WD, Surmeier DJ, Chung JM, et al. Classification of somatosensory tract cells by cluster analysis. In: Schmidt RF, Schaible H-G, Vahle-Hinz C, eds. Fine afferent nerve fibers and pain. Weinheim, Germany: VCH Verlagsgesellschaft, 1987:377–389.
47. Willis WD, Trevino DL, Coulter JD, Maunz RA. Responses of primate spinothalamic tract neurons to natural stimulation of the hindlimb. J Neurophysiol 1974; 37:358–372.
48. Gerhart KD, Yezierski RP, Giesler GJ Jr, Willis WD. Inhibitory receptive fields of primate spinothalamic tract cells. J Neurophysiol 1981; 46:1309–1325.
49. Foreman RD, Schmidt RF, Willis WD. Effects of mechanical and chemical stimulation of fine muscle afferents upon primate spinothalamic tract cells. J Physiol 1979; 286:215–231.
50. Milne RJ, Foreman RD, Giesler GJ Jr, Willis WD. Convergence of cutaneous and pelvic visceral nociceptive inputs onto primate spinothalamic tract cells. Pain 1981; 11:163–183.
51. Ammons WS, Girardot MN, Foreman RD. $T_2$–$T_5$ spinothalamic neurons projecting to medial thalamus with viscerosomatic input. J Neurophysiol 1985; 54:73–89.
52. Blair RW, Weber RN, Foreman RD. Responses of thoracic spinothalamic neurons to intercardiac injection of bradykinin in the monkey. Circ Res 1982; 51:83–94.
53. Foreman RD. Organization of the spinothalamic tract as a relay for cardiopulmonary sympathetic afferent fiber activity, vol. 9. In: Ottoson D, ed. Progress in sensory physiology. New York: Springer-Verlag, 1989, 1–51.
54. Blair RW, Ammons WS, Foreman RD. Responses of thoracic spinothalamic and spinoreticular cells to coronary artery occlusion. J Neurophysiol 1984; 51:636–648.
55. Ammons WS, Blair RW, Foreman RD. Greater splanchnic excitation of primate $T_1$–$T_5$ spinothalamic neurons. J Neurophysiol 1984; 51:592–603.
56. Ammons WS. Characteristics of spinoreticular and spinothalamic neurons with renal input. J Neurophysiol 1987; 58:480–495.
57. Ammons WS. Primate spinothalamic cell responses to urethral occlusion. Brain Res 1989; 496:124–130.
58. Kevetter GA, Haber LH, Yezierski RP, Chung JM, Martin RF, Willis WD. Cells of origin of the spinoreticular tract in the monkey. J Comp Neurol 1982; 207:61–74.
59. Haber LH, Moore BD, Willis WD. Electrophysiological response properties of spinoreticular neurons in the monkey. J Comp Neurol 1982; 207:75–84.
60. Yezierski RP. Spinomesencephalic tract: projections from the lumbosacral spinal cord of the rat, cat, and monkey. J Comp Neurol 1988; 267:131–146.
61. Zhang D, Carlton SM, Sorkin LS, Willis WD. Collaterals of primate spinothalamic tract neurons to the periaqueductal gray. J Comp Neurol 1990; 296:277–290.
62. Yezierski RP, Schwartz RH. Response and receptive-field properties of spinomesencephalic tract cells in the cat. J Neurophysiol 1986; 55:76–96.
63. Yezierski RP, Sorkin LS, Willis WD. Response properties of spinal neurons projecting to midbrain or midbrain thalamus in the monkey. Brain Res 1987; 437:165–170.
64. Craig AD, Kniffki KD. Spinothalamic lumbosacral lamina I cells responsive to skin and muscle stimulation in the cat. J Physiol 1985; 365:197–221.
65. Apkarian AV, Hodge CJ. The primate spinothalamic pathway. II. The cells of origin of the dorsolateral and

ventral spinothalamic pathways. J Comp Neurol 1989; 288:474–492.

66. Hylden JLK, Mayashi M, Dubner R, Bennett GJ. Physiology and morphology of the lamina I spinomesencephalic projection. J Comp Neurol 1986; 247:505–515.
67. Apkarian AV, Hodge CJ. A dorsolateral spinothalamic tract in macaque monkey. Pain 1989; 37:323–333.
68. Craig AD. Ascending lamina I axons in the cat are concentrated in the middle of the lateral funiculus. Neurosci Abstr 1989;15:1190.
69. Cervero F, Iggo A, Molony V. Responses of spinocervical tract neurones to noxious stimulation of the skin. J Physiol 1977; 267:537–558.
70. Boivie J. Thalamic projections from the lateral cervical nucleus in monkey. A degeneration study. Brain Res 1980; 198:12–26.
71. Downie JW, Ferrington DG, Sorkin LS, Willis WD Jr. The primate spinocervicothalamic pathway: responses of cells of the lateral cervical nucleus and spinocervical tract to innocuous and noxious stimuli. J Neurophysiol 1988; 59:861–885.
72. Angaut-Petit D. The dorsal column system. I. Existence of long ascending postsynaptic fibres in the cat's fasciculus gracilis. Exp Brain Res 1975; 22:457–470.
73. Angaut-Petit D. The dorsal column system. II. Functional properties and bulbar relay of the postsynaptic fibres of the cat's fasciculus gracilis. Exp Brain Res 1975; 22:471–493.
74. Rustioni A, Kaufman AB. Identification of cells of origin of non-primary afferents to the dorsal column nuclei of the cat. Exp Brain Res 1977; 27:1–14.
75. Ferrington DG, Downie JW, Willis WD. Primate nucleus gracilis neurons: responses to innocuous and noxious stimuli. J Neurophysiol 1988; 59:886–907.
76. Fields HL, Emson PC, Leigh BK, Gilbert RFT, Iversen LL. Multiple opiate receptors on primary afferent fibres. Nature 1980; 284:351–353.
77. Hamon M, Gallissot MC, Menard F, Gozlan H, Bourgoin S, Verge D. 5-$HT_3$ receptor binding sites are on capsaicin-sensitive fibres in the rat spinal cord. Eur J Pharmacol 1989; 164:315–322.
78. Barber RP, Vaughn JE, Saito K, McLaughlin BJ, Roberts E. Gabaergic terminals are presynaptic to primary afferent terminals in the substantia gelatinosa of the rat spinal cord. Brain Res 1978; 141:35–55.
79. Kuraishi Y, Hirota N, Sato Y, Kaneko S, Satoh M, Takagi H. Noradrenergic inhibition of the release of substance P from the primary afferents in the rabbit spinal dorsal horn. Brain Res 1985; 359:177–182.
80. Jessel TM, Iversen LL. Opiate analgesics inhibit substance P release from rat trigeminal nucleus. Nature 1977; 268:549–551.
81. Aimone LD, Magaman CL, Yaksh TL. Noradrenergic regulation of calcitonin gene-related peptide (CGRP) and substance P (SP) release from rat dorsal horn slices. Neurosci Abstr 1989; 15:549.
82. Solomon RE, Aimone LD, Yaksh TL, Gebhart GF. Release of calcitonin gene-related peptide (CGRP) from rat spinal cord in vitro. Neurosci Abstr 1989; 15:549.
83. Lembeck F, Donnerer J. Opioid control of the primary afferent substance P fibres. Eur J Pharmacol 1985; 114:190–195.
84. Piercey MF, Schroeder LA, Folkers K, Horig XJ. Sensory and motor functions of spinal cord substance P. Science 1981; 214:1361–1362.
85. Post C, Paulsson I. Antinociceptive and neurotoxic actions of substance P analogues in the rat's spinal cord after intrathecal administration. Neurosci Lett 1985; 57:159–164.
86. Vaught JL, Scott R. Species differences in the behavioral toxicity produced by intrathecal substance P antagonists: relationship to analgesia. Life Sci 1987; 40:175–181.
87. Gordh T, Post C, Olsson Y. Evaluation of the toxicity of subarachnoid clonidine, guafacine and substance-P antagonists on rat spinal cord and nerve roots: light and electron microscope observations after chronic intrathecal administration. Anesth Analg 1986; 65:1303–1311.
88. Aanonsen LM, Wilcox GL. Nociceptive action of excitatory amino acids in the mouse: effects of spinally administered opioids, phencyclidine and sigma agonists. J Pharmacol Exp Ther 1987; 243:9–19.
89. Aanonsen LM, Wilcox GL. Muscimol, $\gamma$-aminobutyric $acid_A$ receptors and excitatory amino acids in the mouse spinal cord. J Pharmacol Exp Ther 1989; 248:1034–1038.
90. Huang AS, Wilcox GL. Baclofen, gamma-aminobutyric acid B receptors and substance P in the mouse spinal cord. J Pharm Exp Ther 1989; 248:1026–1033.
91. Willcockson WS, Chung JM, Hori Y, Lee KH, Willis WD. Effects of iontophoretically released amino acids and amines on primate spinothalamic tract cells. J Neurosci 1984; 4:732–740.
92. Engberg I, Ryall RW. The inhibitory action of noradrenaline and other monoamines on spinal neurons. J Physiol 1966; 185:298–322.
93. Zieglgaensberger W, Sutor B. Responses in substantia gelatinosa neurons to putative neurotransmitters in an in vitro preparation of the adult rat spinal cord. Brain Res 1983; 279:316–320.
94. Duggan AW, Johnson SM, Morton CR. Differing distributions of receptors for morphine and met-enkephalinamide in the dorsal horn of the cat. Brain Res 1981; 229:379–387.
95. Headly PM, Duggan AW, Griersmith BT. Selective reduction by noradrenaline and 5-hydroxytryptamine of nociceptive responses of cat dorsal horn neurones. Brain Res 1978; 145:185–189.
96. Carlton SM, Chung JM, Leonard RB, Willis WD. Funicular trajectories of brainstem neurons projecting to the lumbar spinal cord in the monkey (*Macaca fascicularis*): a retrograde labeling study. J Comp Neurol 1985; 241:382–404.
97. Basbaum AI, Zahs K, Lord B, Lakos S. The fiber caliber of 5-HT immunoreactive axons in the dorsolateral funiculus of the spinal cord of the rat and cat. Somatosensory Res 1988; 5:177–185.
98. Stevens RT, Apkarian AV, Hodge CJ. Funicular course of catecholamine fibers innervating the lumbar spinal cord of the cat. Brain Res 1985; 336:243–251.
99. Sorkin LS, Hughes MG. Bilateral release of 5HT into the dorsal horn of the cat lumbar spinal cord elicited by intradermal capsaicin injection. Neurosci Abstr 1989; 15:549.
100. Tyce GM, Yaksh TL. Monoamine release from cat spinal cord by somatic stimuli: an intrinsic modulatory system. J Physiol 1982; 314:513–529.
101. Yaksh TL, Tyce GM. Release of norepinephrine and serotonin in cat spinal cord: direct in vivo evidence for the activation of descending monoamine pathways by somatic stimulation. J Physiol 1980; 77:483–487.
102. Proudfit HK, Yaksh TL. Alterations in nociceptive threshold and morphine-analgesia following the selec-

tive depletion of spinal cord monoamines. Neurosci Abstr 1980; 6:433.

103. Yaksh TL, Wilson PR. Spinal serotonin terminal system mediates antinociception. J Pharmacol Exp Ther 1979; 208:446–453.

104. Bowker RM, Abbott LC, Dilts RP. Peptidergic neurons in the nucleus raphe magnus and the nucleus gigantocellularis: their distributions, interrelationships, and projections to the spinal cord. In: Fields HL, Besson J-M, eds. Progress in brain research. Vol 77. Amsterdam: Elsevier, 1988:95–127.

105. Millhorn DE, Hokfelt T, Seroogy K, Oertel W, Verhofstad AAJ, Wu J-Y. Immunohistochemical evidence for colocalization of $\gamma$-aminobutyric acid and serotonin in neurons of the ventral medulla oblongata projecting to the spinal cord. Brain Res 1987; 410:179–185.

106. Sorkin LS, Willis WD, McAdoo DJ. Release of serotonin following brainstem stimulation: correlation with inhibition of dorsal horn neurons. In: Besson JM, ed. Serotonin and pain. Amsterdam: Elsevier 1990:105–115.

107. Abhold RH, Bowker RM. Descending modulation of dorsal horn biogenic amines as determined by in vivo dialysis. Neurosci Lett 1990; 108:231–236.

108. Proudfit HK. Reversible inactivation of raphe magnus neurons: effects on nociceptive threshold and morphine-induced analgesia. Brain Res 1980; 201:459–464.

109. Schlosberg AJ, Harvey JA. Diurnal changes in serotonin content of frontal pole and pain sensitivity in the rat. Physiol Behav 1978; 20:117.

110. Proudfit HK, Anderson EG. Morphine analgesia: blockade by raphe magnus lesions. Brain Res 1975; 98:612–618.

111. Kellstein DE, Malseed RT, Goldstein FJ. Opioid-monoamine interactions in spinal nociception: evidence for serotonin but not norepinephrine reciprocity. Pain 1988; 34:85–92.

112. Contreras E, Tamaya L. Influence of changes on brain 5-hydroxytryptamine on morphine analgesia. Arch Biol Med Exp 1967; 4:190.

113. Minor BG, Post C, Archer T. Blockade of intrathecal 5-hydroxytryptamine induced antinociception in rats by noradrenaline depletion. Neurosci Lett 1985; 54:39–44.

114. Danysz W, Jonsson G, Minor BG, Post C, Archer T. Spinal and locus coeruleus noradrenergic lesions abolish the analgesic effects of 5-methoxy-*N*, *N*-dimethyltryptamine. Behav Neural Biol 1986; 46:71–86.

115. Archer T, Danysz W, Jonsson G, Minor BG, Post C. 5-methoxy-*N*, *N*-dimethyltryptamine induced analgesia is blocked by $\alpha$-adrenoceptor antagonists in rats. Br J Pharmacol 1986; 89:293–298.

116. Archer T, Jonsson G, Minor BG, Post C. Noradrenergic-serotonergic interactions and nociception in the rat. Eur J Pharmacol 1986; 120:295–307.

117. Westlund KN, Coulter JD. Descending projections of the locus coeruleus and subcoeruleus/medial parabrachial nuclei in monkey: axonal transport studies and dopamine-beta hydroxylase immunocytochemistry. Brain Res Rev 1980; 2:235–264.

118. Crawley JN, Roth RH, Maas JW. Locus coeruleus stimulation increases noradrenergic metabolite levels in rat spinal cord. Brain Res 1979; 166:180–184.

119. Hammond DL, Tyce GM, Yaksh TL. Efflux of 5-hydroxytryptamine and noradrenaline into spinal cord superfusates during stimulation of the rat medulla. J Physiol 1985; 359:151–162.

120. Jones SL, Gebhart GF. Characterization of coeruleospinal inhibition of the nociceptive tail-flick reflex in the rat: mediation by spinal $\alpha_2$-adrenoceptors. Brain Res 1986; 364:315–330.

121. Reddy SVR, Yaksh TL. Spinal noradrenergic terminal system mediates antinociception. Brain Res 1980; 189:391–401.

122. Janss AJ, Jones SL, Gebhart GF. Effect of spinal norepinephrine depletion on descending inhibition of the tail flick reflex from the locus coeruleus and lateral reticular nucleus in the rat. Brain Res 1987; 400:40–52.

123. Sagan J, Proudfit HK. Effect of intrathecally administered noradrenergic antagonists on nociception in the rat. Brain Res 1984; 310:295–301.

124. Bodnar RJ, Ackermann RF, Kelly DD, Glusman M. Elevations in nociceptive thresholds following locus coeruleus lesions. Brain Res Bull 1978; 3:125–130.

125. Heller B, Saavedra JM, Fischer E. Influence of adrenergic blocking agents upon morphine and catecholamine analgesic effect. Experimentia 1968; 24:804–805.

126. Coombs DW. Intraspinal analgesia infusion by implanted pump. Ann NY Acad Sci 1988; 531:108–122.

127. Fielding SJ, Wilker M, Hynes M, Szewczak M, Novick WJ Jr, Lal H. A comparison of clonidine with morphine for antinociceptive and antiwithdrawal actions. J Pharmacol Exp Ther 1978; 207:899–915.

128. Gold MS, Redmond DE, Kleber HD. Clonidine blocks acute opioid withdrawal symptoms. Lancet 1978; 2:599–601.

129. Taiwo YO, Fabian A, Pazoles CJ, Fields HL. Potentiation of morphine antinociception by monoamine reuptake inhibitors in the rat spinal cord. Pain 1985; 21:329–337.

130. Sindrup SH, Gram LF, Brosen K, Eshoj O, Mogensen EF. The selective serotonin reuptake inhibitor paroxetine is effective in the treatment of diabetic neuropathy symptoms. Pain 1990; 42:135–144.

131. Rosenberg PH, Nissinen E, Mannisto PT, Tuomisto L, Heavner JE. Differential effect of tetracaine and bupivacaine on cathecholamine and 5-hydroxytryptamine turnover. Pain 1987; 29:239–245.

132. DeFalque RJ. Potentiation of morphine by intraperitoneal injections of iproniazid in rabbits. Anesth Analg 1965; 44:190–193.

# The Difference between Acute and Chronic Pain

Katherine P. Grichnik, M.D., and F. Michael Ferrante, M.D.

Pain has affected humankind since antiquity (1–3). Experiencing pain has always been a reality, and seeking relief from pain has probably always been a natural response (4). Pain, one of the most complex human experiences, involves physiologic, psychologic, and emotional responses (5). Pain has been divided into acute and chronic phases, originally based on temporal relationships.

Bonica (6) has defined acute pain as "a constellation of unpleasant sensory, perceptual, emotional and mental experiences with associated autonomic, psychological and behavioral responses, provoked by injury, potential injury, or acute disease." The International Association for the Study of Pain defines chronic pain as that pain which persists beyond the expected normal time of healing (7). In practice, this may be less than one month or more than six months. Three months is taken as a convenient point of division. As discussed in this article, the differences between acute and chronic pain are multiple and not just temporal.

Teleologically, acute pain has a definite purpose: It is a response to acute disease or injury and acts as a warning signal. In some situations, acute pain initiates withdrawal and protective reflexes. Chronic pain serves no apparent useful purpose and does not necessarily result from disease or injury.

This brief article reviews the scope of, differences between, and therapeutic modalities for acute and chronic pain.

## Scope

**Acute Pain.** Acute pain is one of the most frequent reasons people seek help from medical practitioners (8). Approximately 64 million people a year suffer injuries, 49 million people a year are afflicted with acute disease, and 23 million people a year undergo operations (8). Of these patients, most are undertreated or suffer from incomplete relief of their pain.

With the implementation of improved forms of pain control in the past decade, much interest has been generated in the management of acute pain. Treatment of acute pain is critical to the total care of the patient. Provision of adequate analgesia allows the patient to feel better by diminishing discomfort, and to function better by interrupting deleterious reflexes. Examples of such reflexes include splinting; bowel and bladder hypomotility; uterine vasoconstriction; myocardial insufficiency; and shock. The alleviation of pain can lead to earlier mobility and ambulation; decreased postoperative complications; increased fetal viability; and decreased mortality. Furthermore, appropriate treatment of acute pain decreases the likelihood of progression to a chronic pain state.

**Chronic Pain.** Chronic pain, a devastating problem, is one of the leading causes of disability in the United States. Bonica estimates that 25%–30% of the populations of industrialized nations are afflicted with chronic pain. One-half to two-thirds are totally disabled for some period of time. Some of these patients are permanently disabled (8).

The consequences of chronic pain include lost work days, excessive hospitalization, unnecessary operations, and overmedication. Although there are no good studies on the economics of chronic pain, it is estimated that chronic pain syndromes cost the American people about $65 billion a year

---

From the Pain Treatment Service, Department of Anesthesia, Brigham and Women's Hospital, Harvard Medical School, Boston, MA. Address reprint requests to F. Michael Ferrante, M.D., Director, Pain Treatment Service, Assistant Professor of Anesthesia, Department of Anesthesia, Brigham and Women's Hospital, Harvard Medical School, Boston, MA 02115.

(8). Further, the social sequelae of chronic pain are equally disruptive, leading to estrangement from families; loss of jobs; dependency on relatives and friends; and suicide (8).

Cancer pain is a special subset of chronic pain. Approximately 3.5 million people in the United States have cancer (4). Traditionally, people with cancer have been undermedicated because of a lack of appreciation of the severity of their pain; unfounded fears of addiction; and overregulation of narcotics (4). Addiction is rarely the result of treating pain, in those with a nonaddicted premorbid personality (9).

## Physiology

Pain perception is a complex experience influenced by emotional, arousal, and thought processes (5). Our knowledge of the physiology of pain as well as of the differences between acute and chronic pain is incomplete, but under intense investigation. (See ruled box for terminology.)

**Acute Pain.** Acute pain is thought to arise from a specific tissue injury or illness. In brief, a mechanical, electrical, or chemical stimulus is detected by nociceptors, which send a neural impulse to the dorsal root ganglion via A delta and C fibers. These afferent fibers enter the dorsal horn of the spinal cord and synapse in distinct laminae of Rexed (I, II, III, V) (5). The spinothalamic system transmits nociceptive signals to more rostral areas, including the thalamus, cortex, reticular activating system, and limbic system. Specific chemical modulators of pain, not reviewed here, include endorphins, enkephalins, and serotonin.

Acute pain elicits reflexes, including skeletal muscle contraction and autonomic (especially sympathetic) nervous system activation (8–11). Tachycardia, increased cardiac output, hypertension, decreased gastric motility, vasoconstriction, increased ventilation, diaphoresis, and mydriasis are characteristically associated with perception of acute pain. These reflexes facilitate escape from the cause of the injury or coping with the injury that has occurred. However, left untreated, these responses become maladaptive and can perpetuate or exacerbate the injury or illness (6, 8).

**Chronic Pain.** Chronic pain is a more complex problem; its physiology is poorly defined. Consequently, one must rely on hypotheses and theories to understand it. Bonica has proposed that chronic pain be divided into syndromes based on peripheral mechanisms, peripheral-central mechanisms, central mechanisms, and psychological mechanisms (8).

Peripheral mechanisms result from persistent noxious stimulation of nociceptors. Sensitization of nociceptors may occur via the release of damaging vasoactive substances from cell degeneration. Pain emanating from skeletal muscle, peptic ulcers, pancreatitis, and arthritis is largely the result of peripheral mechanisms.

Peripheral-central mechanisms involve abnormal function of both the peripheral and central portions of the somatosensory system. Pain is associated with partial or complete lesions of the peripheral nervous system, resulting in loss of descending inhibitory influences and spontaneous firing of regenerated nerve fibers. Examples include causalgia, phantom limb pain, and postherpetic neuralgia.

Central mechanisms are invoked for such conditions as thalamic lesions, tabes dorsalis, and central nervous system diseases. The actual mechanisms are unknown, but may involve loss of descending inhibitory influences.

Psychological mechanisms are divided into four groups: psychosomatic illness, in which emotional stress leads to physiologic reactions; psychogenic causes, in which pain is an associated reaction to anxiety; operant mechanisms, in which the expression of pain produces secondary gain; and psychiatric pain, associated with such illnesses as schizophrenia.

In general, chronic pain does not elicit sympathetic nervous system responses. As painful conditions become prolonged and continuous, autonomic nervous system activity decreases. Persons with chronic pain are also thought to have a depletion of serotonin and endorphins, leading to decreased pain tolerance (6, 8). Often, patients with chronic pain develop vegetative signs, such as loss of appetite, sleep disturbances, loss of libido, and constipation.

## Psychology

The most striking differences between acute and chronic pain are the individual reactions to each one.

**Acute Pain.** The predominant psychological response to acute pain is anxiety (12). This response is further modified by the individual's degree of preparedness as well as the perceived etiology and anticipated duration of the pain. Acute pain usually has a definable cause and can rarely be attributed solely to psychological factors. Further, the emotions provoked by acute pain serve to mobilize a person to react (move, withdraw) to the pain.

**Chronic Pain.** Chronic pain elicits a very different set of psychological responses. One important difference is that the individual cannot give meaning to the pain (12). The pain serves no purpose as a warning signal that would lead to avoidance or treatment. Characteristic responses include despair, fear, and hopelessness. This in turn predisposes to overt depression, hypochondriasis, and somatic focusing. Further, cognitive behavior changes can occur secondary to excess medication.

Chronic pain patients also become enmeshed in the medical community as they go from doctor to doctor and test to test in an effort to have the pain diagnosed and treated. For many patients, the pain becomes the central focus of their lives. The patient is reduced to an existence encompassing only home, pharmacy, and doctors' offices (6). As patients withdraw from society, they lose their jobs, alienate family and friends, and become isolated. Seventy percent of chronic pain patients divorce, and 20% contemplate or commit suicide (13).

## Therapeutics

The therapeutic approaches to acute and chronic pain also differ markedly. In acute pain, therapy is directed at interrupting the nociceptive signal. One therapeutic measure is peripheral neural blockade. Spinal and epidural anesthesia/analgesia interrupts propagation of neural impulses at the level of the spinal cord. These modalities are useful not only for surgical pain control, but also for medical illness.

Traditionally, parenteral narcotics have been the mainstay of acute pain management, but the amount and dosing intervals have been inadequate for complete pain relief (14, 15). Fortunately, this situation is slowly changing as the benefits of adequate analgesia are recognized. Alternative methods of delivery, such as patient-controlled analgesia, are frequently employed.

**Chronic Pain.** Management of chronic pain is more complex. The treatment of choice is prevention. Attempts at preventing chronic pain require early and adequate therapy for acute pain. Further, the treatment of chronic pain must be multidisciplinary, involving specialists in anesthesiology, physical therapy, social services, and surgery.

Specific medical treatment of chronic pain requires an exhaustive search for organic illness that may have initiated or maintained chronic pain. Whether found or not and whether treatable or not, further therapy is often reduced to neural blockade, nonnarcotic medications, electrical modalities, and narcotics (16–18).

Pain Terms

| | |
|---|---|
| **Allodynia** | pain due to nonnoxious stimulus to normal skin |
| **Analgesia** | absence of pain on noxious stimulation; relief of pain without loss of consciousness |
| **Causalgia** | a syndrome of sustained burning pain, allodynia, and hyperpathia after a traumatic nerve lesion, combined with vasomotor and sudomotor dysfunction with trophic changes |
| **Central pain** | pain associated with a lesion of the central nervous system |
| **Dysesthesia** | a spontaneous or evoked unpleasant, abnormal sensation |
| **Hyperalgesia** | increased sensitivity to noxious stimulation |
| **Hyperpathia** | increased reaction to stimulus |
| **Hypalgesia** | decreased sensitivity to noxious stimulation |
| **Neuralgia** | pain in the distribution of a nerve or nerves |
| **Neuritis** | inflammation of a nerve or nerves |
| **Neuropathy** | a disturbance of function or pathologic change in a nerve |
| **Nociceptor** | a receptor preferentially sensitive to a noxious or potentially noxious stimulus |
| **Pain** | an unpleasant sensory and emotional experience associated with actual or potential tissue damage, or described in terms of such damage |
| **Pain threshold** | the least-intense stimulus at which a subject perceives pain |
| **Pain tolerance level** | the greatest stimulus intensity causing pain that a subject is prepared to tolerate |

*Adapted* from Merskey H. Pain terms . . . recommended by the IASP Subcommittee on Taxonomy. Pain 1979; 6:249–252.

Frequently used blockade techniques include epidural steroid injections for low back pain and efferent sympathetic interruption for reflex sympathetic dystrophies. Nonnarcotic medications provide adjunctive therapy for specific kinds of pain. Tricyclic antidepressants can be used for cancer pain; anticonvulsants for neuropathic pain; and nonsteroidal anti-inflammatories for metastatic bone pain. Electrical modalities include transcutaneous nerve stimulation and dorsal column stimulation.

The judicious use of narcotics has a role in chronic pain management, especially for the cancer patient. Narcotics can be delivered to the patient through several routes, including the oral, subcutaneous, epidural, and subarachnoid. Cancer patients often require extraordinary amounts of narcotics. Narcotics should not be withheld for fear of addiction or respiratory depression.

## Summary

Acute and chronic pain are different clinical entities. Acute pain is provoked by a specific disease or injury, serves a useful biologic purpose, is associated with skeletal muscle spasm and sympathetic nervous system activation, and is self-limited. Chronic pain, in contrast, may be considered a disease state (8). It is pain that outlasts the normal time of healing, if associated with a disease or injury. Chronic pain may arise from psychological states, serves no biologic purpose, and has no recognizable end-point. Both acute and chronic pain are an enormous problem in the United States, costing 650 million lost workdays and $65 billion a year (8). The therapy of acute pain is aimed at treating the underlying cause and interrupting the nociceptive signals. The therapy of chronic pain must rely on a multidisciplinary approach and should involve more than one therapeutic modality.

### References

1. Fairley P. The conquest of pain. New York: Scribner's, 1978:19–54.
2. Seeman B. Man against pain: 3,000 years of effort to understand and relieve physical suffering. Philadelphia: Chilton, 1962:1–13.
3. Smolkin MT. Understanding pain: interpretation and philosophy. Malabar, FL: Robert E Krieger, 1989:1–11.
4. Raj PP. Pain relief: fact or fancy? Reg Anesth 1990; 15(4):157–169.
5. Chapman CR. Psychological factors in postoperative pain. In: Covino BG, Smith G, eds. Acute pain. London: Butterworths, 1985:22–41.
6. Bonica JJ. Important clinical aspects of acute and chronic pain. In: Beers RF, Bassett EG, eds. Mechanisms of pain and analgesic compounds. New York: Raven Press, 1979:15–29.
7. Merskey H, ed. International Association for the Study of Pain: classification of chronic pain; description of chronic pain syndromes and definitions of pain states. Pain 1986; (Suppl 3):S1.
8. Bonica JJ. The management of pain, 2nd ed. Philadelphia: Lea & Febiger, 1990:159–197.
9. Kornfeld DS. Psychiatric factors in the management of pain. In: Levitan SJ, Berkowitz HL, eds. Pain research and treatment. Washington, DC: American Psychiatric Press, 1985:53–57.
10. Sternbach RA. Clinical aspects of pain. In: Sternbach RA, ed. The psychology of pain. New York: Raven Press, 1986:223–224.
11. Bonica JJ. Neurophysiologic and pathologic aspects of acute and chronic pain. Arch Surg 1977; 112:750–761.
12. Sternbach RA. Pain patient: traits and treatment. New York: Academic Press, 1974:5–10.
13. Hendler N. Cited by Neff PN. Med World News 1976; 17:54–60.
14. Marks RM, Sachar EJ. Undertreatment of medical patients with narcotic analgesics. Ann Intern Med 1973; 78(2):173–181.
15. Smith G. Management of postoperative pain. Can J Anaesth 1989; 36(3):S1–S4.
16. Budd K. Recent advances in the treatment of chronic pain. Br J Anaesth 1989; 63(2):207–212.
17. Raj PP. The management of chronic pain by the anesthesiologist. In: Zorab JSM, ed. Lectures in anesthesiology. Oxford: Blackwell Scientific, 1986:39–51.
18. Wyant GM. The nature and principles of the treatment of chronic pain. In: Zorab JSM, ed. Lectures in anesthesiology. Oxford: Blackwell Scientific, 1988:51–63.
19. Merskey H. Pain terms: a list with definitions and notes on usage, recommended by the IASP Subcommittee on Taxonomy. Pain 1979; 6:249–252.

# Nonnarcotic Analgesics and Tricyclic Antidepressants for the Treatment of Chronic Nonmalignant Pain

DAVID M. RICHLIN, M.D.

PAIN is "an unpleasant sensory and emotional experience associated with actual or potential tissue damage, or described in terms of such damage" (1). This definition encompasses three fundamental components of the pain experience. Pain is unpleasant. It is nearly always perceived by the sufferer as a sign of a physical problem. It is an emotional experience that may have a significant impact on mood and behavior.

Chronic nonmalignant pain is characterized by unanticipated persistence (usually for more than three months) and intractability. It may be difficult to establish correlation of organicity with the pain complaint and apparent suffering. When this pattern is accompanied by persistent care-seeking and invalidism, it is referred to as the "chronic pain syndrome." Patients with chronic pain commonly receive various oral medications, including nonsteroidal anti-inflammatory drugs (NSAIDs), opioid analgesics, sedatives, and anxiolytics. A history of multiple trials of various medications, failure to achieve sustained pain relief, and a pattern of disorganized, illogical, excessive, and inappropriate drug ingestion, even habituation to potent narcotics are all characteristic of patients with chronic pain syndrome. The persistence of this syndrome and its reinforcement by a succession of failed therapeutic interventions engender secondary disorders characterized by dependence, depression, frustration, helplessness, drug side effects, toxicity, and addiction.

Whatever the role of organic causality in chronic nonmalignant pain, correction of the consequences of ineffective or inappropriate drug treatment and the maladaptive behavioral response are integral to effective treatment.

## Patient Evaluation

Rational drug therapy with nonnarcotic analgesics, NSAIDs, and antidepressants in patients with chronic nonmalignant pain requires careful patient assessment and identification of physical, emotional, behavioral, and environmental factors that contribute to the chronic pain state.

**Prior and Ongoing Drug Use.** The comprehensive evaluation of the patient must include a review of all drugs used, past and present; their doses and patterns of ingestion, and the effects of drugs on the patient's pain experience and level of activity. Consideration must be given to coincident use of medication for non–pain-related medical problems. Medication for chronic disorders may have significant pharmacokinetic and pharmacodynamic effects on drugs for pain control, and vice versa, particularly if these drugs are highly protein bound, as are warfarin and phenytoin. Almost all the NSAIDs are up to 90% protein bound, so coadministered medication may cause significant, potentially hazardous alteration in clinical efficacy, side effects, and toxicity.

## Principles of Proper Drug Use To Treat Chronic Nonmalignant Pain

The management plan must correct any maladaptive drug use by taper withdrawal of inappropriate narcotics, benzodiazepines, and seda-

---

From the Department of Anesthesiology, Mount Sinai School of Medicine (CUNY), New York, New York. Address reprint requests to David M. Richlin, M.D., Department of Anesthesiology, Box 1192, Mount Sinai School of Medicine (CUNY), One Gustave L. Levy Place, New York, NY 10029.

tives (2). Withdrawal of these agents in patients with the chronic pain syndrome may account for up to 50% of the improvement achieved in their condition (3).

For the management plan, the physician may decide that an oral analgesic, NSAID, or tricyclic antidepressant is indicated. Regardless of the specific drug or drugs employed, certain general guidelines apply.

First, administer drugs that are most appropriate to the organic disorder deemed to be the major cause of pain. NSAID drugs are potent anti-inflammatory agents and achieve significant concentrations in synovial fluid as well as plasma. Consequently, they are most useful in chronic arthritic disorders and other pain syndromes characterized by inflammation. The tricyclic antidepressants (TCAs) are indicated in endogenous depression, although they may act to diminish pain perception in patients without depression.

Second, when two or more agents may have similar efficacy, introduce the one least likely to cause side effects or toxicity. Nonnarcotic analgesics, NSAIDs, and TCAs may be associated with a 30% failure rate due to unacceptable side effects or toxicity (4). The most common side effect of aspirin and NSAIDs is gastric irritation, but choice of drug and mode of administration can significantly affect incidence and severity of gastric complaints. TCAs are associated with anticholinergic effects, which may include dry mouth, blurred vision, urinary retention, and drowsiness (5). However, because their clinical effectiveness is not impaired by once-daily use (6), administering a TCA in a single bedtime dose allows the drug to act as a sleep aid.

**Drug Combinations.** When administering both NSAIDs and TCAs simultaneously, it is not appropriate to use more than one agent of each category. A reasonable trial period is necessary before one can assess clinical efficacy in a particular patient. Seven to ten days are required with most NSAIDs and one to four weeks are needed for the TCAs (7). One must instruct the patient (and spouse) what to expect from the drug and not abandon therapy until a thorough and appropriate trial has been completed. Prior to substituting agents, withdraw the previous NSAID or TCA. Before switching from a monoamine inhibitor to a TCA, at least two weeks must intervene between cessation of the MAO inhibitor and commencing the TCA (8). Any switch from an MAO inhibitor to a TCA should take place under the medical supervision of a psychiatrist experienced in the use of these agents for severe depression.

**Fixed-Interval Dosage.** Nonnarcotic analgesics, NSAIDs, and TCAs must be prescribed and ingested on a fixed time schedule, with a dosage interval appropriate to the drug's pharmacokinetic profile. These agents should never be used sporadically or prescribed "as needed" in chronic pain patients, because this pattern will fail to achieve sustained therapeutic levels of drug in plasma. Only fixed-time dosage can produce steady-state plasma levels in a therapeutically effective range. Although the dosage interval should be fixed, the dosage should be individualized, balancing optimal response against minimal side effects.

**Individual Responses.** It is characteristic of the NSAIDs that an individual patient may have a good response to one agent, while failing to improve on another. These different responses may be due to genetic variations among patients (4). Only by trying each drug sequentially, carefully instructing the patient in the appropriate dose and fixed dosage interval, and maintaining the drug for a sufficient trial period can one determine which NSAID will provide optimal analgesia. Failure of one NSAID to afford satisfactory pain control does not mean that a different NSAID would not be effective.

**Advantages and Expectations.** NSAIDs, aspirin, and acetaminophen have particular advantages for patients with chronic nonmalignant pain. They lack sedative, dysphoric, or euphoric effects. They have either no effects on the central nervous system or central effects that enhance analgesia without affecting mood, alertness, or cognition. They are neither addictive nor associated with tolerance or tachyphylaxis. Therefore, aspirin, acetaminophen, and the NSAIDs do not act to cause or reinforce maladaptive patterns of drug use, as may occur with narcotics and benzodiazepines.

All NSAIDs, as well as aspirin and acetaminophen, exhibit a ceiling effect, an optimal dose beyond which additional medication is not associated with improved analgesia. Further, in some patients these agents may have no efficacy or produce only moderate pain reduction and improvement in function. Physicians should inform the patient about reasonable expectations from use of these agents. Both parties must be alert to subtle or moderate improvements in pain and function over time, while not exceeding maximal recommended doses. These considerations reinforce the importance of combining NSAIDs with other pharmacologic, behavioral, cognitive, and physical conditioning therapies in a comprehensive plan for chronic pain treatment.

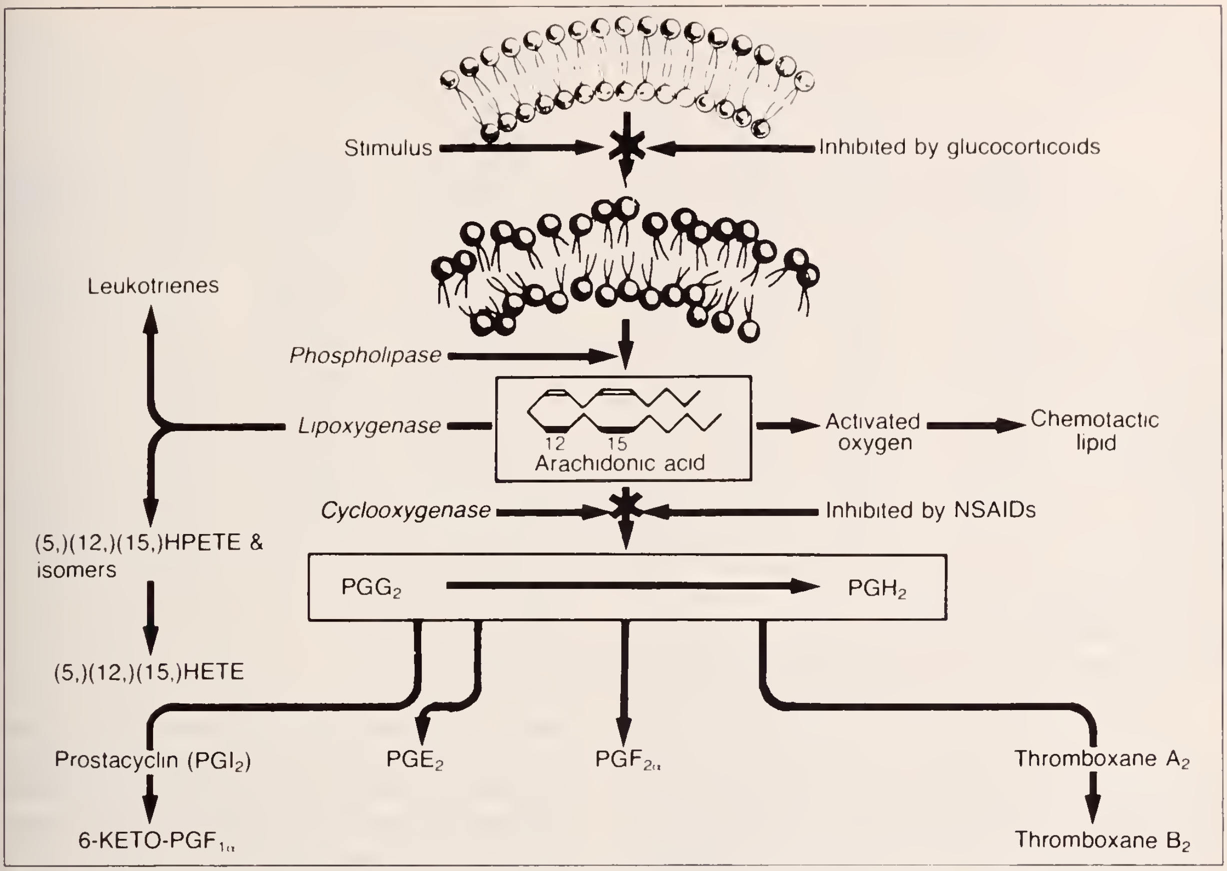


Fig. Arachidonic acid cascade. Bilayer membrane (at top), disrupted by stimulus, makes available in-facing lipid components, which can then be digested by phospholipase to form arachidonic acid. This, in turn, can be acted on by cyclooxygenase and lipoxygenase. NSAIDs inhibit cyclooxygenase strongly and lipoxygenase weakly, if at all. Effect on cyclooxygenase prevents formation of endoperoxides prostaglandins $G_2$ ($PGG_2$) and $H_2$ ($PGH_2$), precluding formation of stable prostaglandins (bottom). Endoperoxides are probable direct mediators of inflammation. Prostaglandins $I_2$ ($PGI_2$) and $E_2$ ($PGE_2$) act with histamine and bradykinin to cause pain. HPETE, hydroperoxyeicosatetraenoic acid; HETE, 12-hydroxyeicosatetraenoic acid. *Reprinted* with permission from Kantor TG (ref. 4).

**Sleep Disturbances.** Sleep disturbances are a major complication of chronic pain problems. Consequently, an important goal of any maintenance program is improvement in sleep. Disruption of the sleep cycle is a common consequence of narcotic and benzodiazepine use, but has not been reported to be a significant problem with the nonnarcotic analgesics. Though pain may prevent or disrupt sleep, and relief of pain may improve sleep, the effects of nonnarcotic analgesics on sleep are not clearly understood. A recent study in arthritic rats demonstrated that aspirin and acetaminophen had opposite effects on sleep. Aspirin increased wakefulness and decreased all sleep stages during the hours of maximal sleep in the rat. These effects were seen in conjunction with significant and equivalent antinociceptive effects in both groups (9). Similarly, in animal studies indomethacin has been shown to decrease sleep (10, 11). These results suggest that there may be a link between prostaglandin synthesis inhibition, particularly the inhibition of cyclooxygenase—not shared by acetaminophen—and disruption of sleep, which overrides the improvement in sleep resulting from analgesia.

## Oral Analgesics: Aspirin and Acetaminophen

Aspirin and acetaminophen are the two most widely used oral analgesics. Acetaminophen differs from aspirin and the NSAIDs in that it appears to have no effect on the enzyme cyclooxygenase, whereas the latter agents affect prostaglandin synthesis by inhibition of cyclooxygenase (Fig.). Inhibition of cyclooxygenase appears to be directly related to the reduction of inflammation, a property not shared by acetaminophen. Aspirin's actual effect on this enzyme is to destroy it, by donation of an acetyl group. One may speculate that this more pronounced effect on cyclooxygenase may correlate with the more profound irritation of gastric mucosa and inhibition of platelet function associated with aspirin compared with the NSAIDs.

TABLE

*Pharmacokinetic Properties of Selected Nonsteroidal Anti-inflammatory Drugs*

| Drug | Usual daily dose* (mg) | Usual dosing interval* (hr) | Time to peak* (hr) | Elimination half-life*† (hr) | Volume of distribution*† (L/kg) | Total clearance‡ (L/kg/hr) | Extent of protein binding† (%) | Fraction metabolized* |
|---|---|---|---|---|---|---|---|---|
| Alclofenac | 300 | 8 | 1–4 | 1.5–5.5 | 0.10 | 0.01–0.05 | 90–99 | 0.5–0.9 |
| Apazone | 1200 | 6 | 3–6 | 10–15 | 0.16 | 0.007–0.01 | 95 | 0.38 |
| Diclofenac | 100 | 8 | 1–3 | 1–2 | 0.12 | 0.04–0.08 | 99 | 0.99 |
| Fenbufen | 600–1,000 | 12 | 1–2 | 10 | 2–4 | 0.14–0.25 | . . . | 0.96 |
| Fenclofenac | 900–1,200 | 12 | 3–4 | 20–38 | 0.2–0.25 | .004–.007 | 96 | |
| Fenoprofen | 2,400 | 6–8 | 1–2 | 2–3 | 0.10 | 0.02–0.04 | 99 | 0.95 |
| Flurbiprofen | 150–300 | 8 | 1–2 | 3–4 | 0.10 | 0.03–0.04 | 99 | 0.85 |
| Ibuprofen | 1,600 | 6–8 | 0.5–1.5 | 2–2.5 | 0.14 | 0.04–0.05 | 99 | 0.99 |
| Indomethacin | 75–100 | 6–12 | 1–2 | 6 | 0.12 | 0.014 | 92–99 | 0.895 |
| Ketoprofen | 200 | 6–8 | 0.5–2 | 1.5 | 0.11 | 0.07 | 94 | 0.99 |
| Naproxen | 500–750 | 12 | 1–2 | 12–15 | 0.10 | .005–.006 | 98–99 | 0.99 |
| Phenylbutazone | 300–400 | 6–8 | 2 | 50–100 | 0.17 | .001–.002 | 98 | 0.99 |
| Sulindac | 400 | 12 | 1 | 7 | . . . | . . . | 93 | 0.93 |

* Data from Verbeeck et al. (ref. 25).

† Data from Heel RC, Avery GS (ref. 26).

‡ Estimated from volume of distribution and half-life according to: $Cl = 0.693 \times VD/t_{1/2}$.

Aspirin is still considered the best all-purpose nonnarcotic analgesic (12), but its use in chronic pain is constrained by its propensity to cause gastrointestinal bleeding and impairment of platelet function. The gastrointestinal effects can be reduced by ingestion after meals or administration of enteric-coated preparations.

Acetaminophen is equianalgesic with aspirin but does not cause gastrointestinal bleeding or platelet dysfunction, as does aspirin. Acetaminophen has anti-inflammatory effects and is indicated when evidence of inflammation is lacking, or in patients for whom aspirin and NSAIDs are either contraindicated or poorly tolerated. Although risk of renal dysfunction must be considered when using any prostaglandin inhibitor, this risk is limited when acetaminophen is employed (13).

Acetaminophen may induce life-threatening hepatic toxicity and necrosis when massive overdose (10 g or more) is ingested. The threshold for hepatic toxicity may be significantly lowered in alcoholism, or during coadministration of drugs that induce hepatic microsomal enzymes (14). The occurrence of liver damage in instances of acetaminophen overdose correlates with drug plasma concentration and plasma half-life in the affected patients. Severe damage will ensue in 90% of patients if plasma acetaminophen concentration exceeds 300 μg/mL at 4 hr or 45 μg/mL at 15 hr postingestion. Additionally, if the plasma half-life exceeds 12 hr, coma is likely; a half-life greater than 4 hr predicts hepatic necrosis. Outcome is improved by early diagnosis and treatment. Gastric lavage (within 4 hr of ingestion), treatment with oral or intravenous acetylcysteine (to replenish stores of glutathione), and supportive care are recommended (15).

Acetaminophen absorption correlates with rate of gastric emptying, and its absorption may be enhanced by the coadministration of metochlorpropamide (16). Unlike aspirin and the NSAIDs, acetaminophen is not highly protein bound (25%).

## Nonsteroidal Anti-Inflammatory Drugs

A large number of NSAIDs have been synthesized (Table); approximately 16 of these agents are in clinical use in the United States. All inhibit cyclooxygenase, thereby affecting or impairing the synthesis of various prostaglandins (Fig.). The synthesis of prostaglandins in peripheral sites where injury or inflammation occur appears to be significant in initiating nociception. The enzyme cyclooxygenase converts arachidonic acid to endoperoxides (prostaglandin G2 and H2). These prostaglandins mediate capillary dilatation and leak, which in turn promote inflammation. The unstable endoperoxides lead to the production of prostaglandin E2 and I2 (prostacycline), which sensitize nociceptors.

Although the analgesic properties of NSAIDs are attributed primarily to their demonstrated anti-inflammatory effects (17), the extent of prostaglandin inhibition does not correlate with NSAID potency or efficacy. Additional, as yet unelucidated factors may contribute to pain relief.

NSAIDs share a number of properties. They possess comparable efficacy, though, as noted, a particular patient may respond to one drug and not another. They are all rapidly and nearly completely absorbed from the gastrointestinal mucosa, and all have a small volume of distribution consistent with their high rate of protein binding. Consequently, a small change in protein binding or concomitant administration of other bound drugs may significantly alter their effect.

**Toxicity and Side Effects.** All NSAIDs may produce gastrointestinal intolerance, their most prominent side effect. As many as 30% of patients who might benefit from these agents may have to discontinue them because of gastrointestinal upset. This occurs because prostaglandins present in gastrointestinal mucosa help regulate gastric mucus and acid production. These processes are impeded by the inhibition of prostaglandin synthesis induced by NSAIDs and aspirin, causing decreased mucus, increased acid, and a decrease in mucosa blood supply (4). The degree to which this mechanism may cause clinical symptoms in individual patients is idiosyncratic.

Recently, a prostaglandin analog, misoprostol, has been released for general clinical use. This agent acts to suppress gastric acid secretion and heal gastric ulcers. It is used primarily to prevent gastric ulceration from long-term use of NSAIDs and aspirin. The recommended dose is 200 μg four times a day. It is not yet known whether coadministration of this agent will reduce the NSAID treatment failure rate due to gastrointestinal toxicity in patients with chronic nonmalignant pain (18).

Individuals with marginal or impaired renal function, particularly the elderly, should use NSAIDs with caution, since NSAIDs may interfere with the effects of prostaglandins in regulating renal physiology. In these instances, close medical supervision and periodic monitoring of renal function are indicated. NSAIDs are contraindicated for patients with renal failure who have chronic nonmalignant pain without a significant inflammatory component.

The high incidence of NSAID toxicity and the possible need to administer additional drugs to suppress toxicity will increase the cost of therapy and may also lead to misinterpretation of prescribing instructions and poor compliance. Use of time-release preparations (19), or drugs with long plasma half-lives, where steady-state therapeutic blood levels may be achieved with twice-daily or even once-daily dosage are advantageous in chronic pain patients, in whom long-term therapy, coadministration of other drugs, and problems with compliance are common.

**Phenylbutazone, Indomethacin, and Sulindac.** Phenylbutazone and indomethacin are the most toxic of the NSAIDs, and long-term use should be discouraged. These agents are most effective in the suppression of acute gout or acute inflammatory arthritis. Phenylbutazone has been associated with bone marrow depression and aplastic anemia (20). Indomethacin is associated with a high incidence of gastrointestinal upset, and in debilitated patients may induce or aggravate neuropsychiatric disturbances (20, 21). Surprisingly, coadministration of indomethacin and warfarin does not result in significant alterations of effects of either drug (22). However, gastrointestinal bleeding caused by indomethacin can be exacerbated by warfarin. Consequently, use of indomethacin in patients requiring warfarin is not recommended.

A strategy that may lower incidence of NSAID side effects and toxicity is to employ a drug that is not a potent prostaglandin inhibitor in the ingested form, but whose metabolite exerts its major therapeutic effect. Sulindac, although related chemically to indomethacin, is apparently better tolerated because the ingested form is an inactive prodrug that is converted by liver microsomal enzymes to sulindac disulfide, which is the active anti-inflammatory agent. Because this metabolite has a long elimination half-life, twice-daily dosing produces steady-state therapeutic blood levels.

**Propionic Acids.** The two most widely used drugs of the propionic acid grup of NSAIDs are ibuprofen and naproxen. Ibuprofen may be most useful in mild cases of inflammation and degenerative arthritis. It is not superior to aspirin in its anti-inflammatory action, but it may have significantly lower toxicity in selected patients. The extent of absorption is not reduced by ingestion with food, and dosing with food may further decrease gastrointestinal upset. Ibuprofen has a relatively short elimination half-life—approximately 1.8 to 2 hr—which necessitates frequent dosing.

Naproxen's elimination half-life is 13 hr; the usual recommended dose is 0.5–1.0 g/day in two or three divided doses. In common with nearly all the NSAIDs, naproxen has greater than 90% affinity for plasma proteins and by displacement has potential to increase the effects of anticoagulants and phenytoin.

When a naproxen sustained-release preparation was compared with the standard, the sustained release form was absorbed at a slower rate but to equal extent. In a double-blind study of patients with osteoarthritis, the 1 g sustained-release form demonstrated efficacy equal to 500 mg twice daily of the standard preparation while achieving a significant reduction in gastrointestinal side effects (19). In patients who have chronic pain, once-daily dosing in the evening confers several advantages. Fluctuations in plasma levels are minimized and plasma concentrations in the morning on arising (when pain and stiffness are often maximal) are higher than those achieved with the standard preparation. However, a sustained-release preparation of naproxen is not currently available for general clinical use.

**Phenylacetic Acid Derivatives.** A newer class of NSAIDs is the phenylacetic acid derivatives, which include diclofenac, fenclofenac, and alclofenac. These agents are comparable to the propionic acid derivatives in efficacy and side effects. They are rapidly absorbed following oral administration and achieve peak plasma concentrations in 2 hr. However, diclofenac's oral bioavailability is only 60%. In common with most of the other NSAIDs, the phenylacetic acids are highly protein bound and have a small volume of distribution. Coadministration of diclofenac and aspirin results in an increase in the apparent volume of distribution and an increase in total plasma clearance for diclofenac, secondary to aspirin's competition for plasma binding sites.

Caution must be exercised when diclofenac is contemplated for use in patients taking lithium. Diclofenac was found to decrease lithium renal clearance by 23% and to increase lithium plasma concentrations by a corresponding 26%, thereby increasing the risk of lithium intoxication (23).

## Tricyclic Antidepressants

Chronic nonmalignant pain is often characterized by complaints of pain, suffering, and disability that are either disproportionate to any organic disorder or without an evident physical disorder or lesion. This dilemma of chronic pain has instigated the development of multidisciplinary pain management programs that employ psychological techniques for diagnosis and therapy. It is advisable that comprehensive evaluation of the patient with chronic nonmalignant pain include psychological evaluation by both interview techniques and standardized testing.

The presence of multiple somatic complaints and emotional lability may reflect an overt or underlying depression. Depression in the context of a chronic pain disorder may be reactive to the apparent physical, emotional, and situational stress induced by pain, a consequence of long-term use of narcotic and sedative drugs (including an inappropriate and ineffective pattern of drug ingestion—see above), or an endogenous coexisting disorder that is more a cause than a result of the pain.

Tricyclic antidepressant drugs (TCAs) are clearly indicated in patients with endogenous depression (7). It has been observed that when depressed patients who also suffer from pain are treated with these agents, their pain symptoms improve or resolve as their depression lifts. This has led to widespread use of these agents to treat all patients with chronic pain and depression, even when the diagnosis of depression is equivocal or based on weak criteria. Although a small number of chronic pain patients will be severely depressed, a much larger number will have depressive symptoms of a lesser degree (24).

The successful use of tricyclic antidepressants in depressed pain patients led to multiple trials of these agents in patients with chronic pain whose pain could not be attributed to depression (7). Approximately 50% of these patients reported at least partial relief.

The mechanism by which TCAs induce the reduction of chronic pain is not clear. In depressed patients, the relief of the depression may be the primary mode. In patients with severely disturbed sleep, pain relief may ensue when sleep is improved with a dosage of TCA possessing sedative properties at bedtime. However, although we acknowledge that these mechanisms may play a significant role in pain relief, there remains a strong inference that TCAs may function as analgesics. Amitriptyline is the agent used most often in chronic pain (21).

Tricyclic antidepressants act to improve depression by blocking the reuptake of serotonin and norepinephrine. Two common observations indicate that the relief of pain achieved with TCAs is not dependent on this effect. First, these agents are often effective in central pain (postherpetic neuralgia, thalamic syndrome, denervation dysesthesia) at doses far less than those used for endogenous depression. Second, patients often report significant improvement in pain after only a few days of TCA use, whereas improvement in depression often requires one to four weeks of regular use.

TCAs are highly protein bound, primarily to alpha-1 acid glycoprotein. They possess large volumes of distribution, long plasma half-lives, and are eliminated by extensive hepatic metabolism. TCAs have markedly different rates of elimination from one patient to another. There is poor correlation between dose administered, drug level in plasma, clinical effect, and side effects. Therefore, the physician should introduce these drugs at a low dose, gradually increasing the dose over time, until the desired clinical effect is achieved or toxicity ensues. Increases in dosage should be carried out every three to seven days; for elderly patients, the time interval should be longer.

Amitriptyline is the drug of first choice in pain patients (21), particularly those without overt depression. There are no clear guides to the clinical choice of other TCAs. Like the NSAIDs, it may be necessary to try each drug in turn, until a satisfactory therapeutic response without intolerable side effects is achieved. It is recommended that amitriptyline be administered to chronic pain patients as a single daily dose at bedtime.

## Summary

Chronic nonmalignant pain is often characterized by multiple treatment failures, a pattern of maladaptive behavior, and depression. Often there is a history of inappropriate and excessive use of medications for pain. Prior and ongoing use of narcotics and sedatives acts to compound and aggravate the chronic pain syndrome. A first step in treatment is controlled withdrawal of these agents.

Nonnarcotic analgesics, NSAIDs, and tricyclic antidepressants are commonly employed in patients with chronic pain. Effective use of these agents requires understanding of their pharmacokinetic and pharmacodynamic properties. Use of a fixed-time schedule is necessary to achieve an effective, sustained therapeutic response. Careful patient education and monitoring for side effects and toxicity are necessary, particularly in the elderly and patients with coexisting medical disorders. Incidence of side effects and toxicity may be reduced by choice of drug and modification of dosing regimen.

Nonnarcotic analgesics, TCAs, and NSAIDs are seldom effective by themselves in resolving the pain and distress of patients with chronic nonmalignant pain. This is particularly true when maladaptive behavior coexists. A comprehensive multimodal pain management program encompassing additional pain-relieving strategies and behavior-modifying techniques should be considered and utilized in conjunction with medication.

### References

1. Merskey H. Pain terms; a list with definitions and notes on usage, recommended by the IASP Subcommittee on Taxonomy. Pain 1979; 6:249–252.
2. Stimmel B. Pain, analgesia, and addiction. New York: Raven Press, 1983:119–125.
3. Black RG. The chronic pain syndrome. Surg Clin North Am 1975; 55:999–1011.
4. Kantor TG. Current modalities in arthritic disease. Am J Medicine 1987; 83 (suppl 4B):2–5.
5. Blackwell B, Stefopoulos A, Eudus P. Anticholinergic activity of two tricyclic antidepressants. Am J Psychiatry 1978; 135:722–724.
6. Weise CC, Stein MK, Pereira-Ogan J, et al. Amitriptyline once daily versus three times daily in depressed outpatients. Arch Gen Psychiatry 1980; 37:555–560.
7. Oxman T, Denson DD. Antidepressants and adjunctive psychotropic drugs. In: Raj PP, ed. Practical management of pain. Chicago: Year Book Medical Publishers, 1986:528–532.
8. Baldessarini RJ. Drugs and the treatment of psychiatric disorders. In: Goodman and Gilman's the pharmacological basis of therapeutics. 8th ed. New York: Pergamon Press, 1990:417.
9. Landis CA. Robinson DR, Helms C, Levine JD. Differential effects of acetylsalicylic acid and acetaminophen on sleep abnormalities in a rat chronic pain model. Brain Res 1989; 488:195–201.
10. Krueger JM, Pappenheimer JR, Karnovsky ML. Sleep-promoting effects of muramyl peptides. Proc Natl Acad Sci USA 1982; 79:6102–6106.
11. Naito K, Ueno R, Hayaishi O, Honda K, Inoue S. Supression of sleep by prostaglandin synthesis inhibitors in unrestrained rats. Sleep Res 1987; 16:64.
12. Nash T, Swerdlow M, Metha M. Drug control of pain. In: Anderson S, Bond M, Metha M, Swerdlow M, eds. Chronic noncancer pain. Lancaster, UK: MTP Press, 1987, 147.
13. Koch-Weser J. Acetaminophen. N Engl J Med 1976; 295:1297–1300.
14. Mitchell JR, Jollow DJ, Gillette JR, et al. Acetaminophen induced hepatic injury: protective role of glutathone in man and rationale for therapy. Clin Pharmacol Ther 1974; 16:676–684.
15. Insel PA. Analgesic-antipyretics and anti-inflammatory agents: drugs employed in the treatment of rheumatoid arthritis and gout. In: Goodman and Gilman's the pharmacological basis of therapeutics. 8th ed. New York: Pergamon Press, 1990:658.
16. Hayes AH. Therapeutic implications of drug interactions with acetaminophen and aspirin. Arch Intern Med 1981; 141:301–304.
17. Crepax P, Silvestrini B. Experimental evaluation in laboratory animals of anti-inflammatory analgesic drugs. Arch Ital Biol 1963; 101:444–457.
18. Monk JP, Clissold SP. Misoprostol: a preliminary review of its pharmacodynamic and pharmacokinetics properties, and therapeutic efficacy in the treatment of peptic ulcer disease. Drugs 1987; 33:1–30.

19. Kelly JG, Kinney CD, Devane JG, Mulligan S, Colgan BV. Pharmacokinetics properties and clinical efficacy of once-daily sustained-release naproxen. Eur J Clin Pharmacol 1989; 36:383–388.
20. Williams NE. The role of drug therapy. In: Lipton S, ed. Persistent pain, modern methods of treatment. New York: Grune & Stratton, 1977:237–252.
21. Twycross RG. The relief of pain in far-advanced cancer. Reg Anesth 1980; 5(3):2–11.
22. Helleberg L. Clinical pharmacokinetics of indomethacin. Clin Pharmacokinet 1981; 6:245–258.
23. Denson DD, Mather LE. Nonsteroidal anti-inflammatory agents. In: Raj PP, ed. Practical management of pain. Chicago: Year Book Medical Publishers, 1986:527.
24. Merskey H. Psychological and psychiatric methods in the treatment of pain. In: Anderson S, Bond M, Metha M, Swerdlow M, eds. Chronic non-cancer pain. Lancaster, UK: MTP Press, 1987:184–185.
25. Verbeeck RK, Black JL, Loewen GR. Clinical pharmacokinetics of non-steroidal anti-inflammatory drugs. Clin Pharmacokinet 1983; 8:297–331.
26. Heel RC, Avery GS. Appendix A: Drug data information. In: Avery GS, ed. Drug treatment. New York: Adis Press, 1980:1213.

# Temporomandibular Disorders and Facial Pain

ANDREW KAPLAN, D.M.D., DANIEL BUCHBINDER, D.M.D., AND SUZANNE McCORMICK, M.S., D.D.S.

TEMPOROMANDIBULAR disorders are a group of problems that affect the temporomandibular joints and the muscles of mastication, causing facial pain. The existence of these problems has been a subject of much controversy, as has been their diagnosis and treatment. Today, many dental-school and hospital-based clinics exist specifically to treat this group of maladies, and several refereed journals are totally or partially dedicated to the topic. In addition, at least 20 textbooks have now been published on the subject.

In the past, the failure of the medical and the dental community to recognize these disorders led to many patients going undiagnosed or misdiagnosed. Today, these disorders are in vogue, and the problem has reversed to one of overdiagnosis. Physicians and dentists must be familiar with the pathologic states that affect the temporomandibular joint and its associated muscles so that these abnormalities may be included in the differential diagnosis of patients with complaints of head and neck pain.

The goal of this paper is to present an overview of the pathologic states collectively referred to as temporomandibular disorders. This information will assist practitioners in diagnosis of patients with head and neck pain. In addition, the accepted and most commonly used treatment alternatives are described.

**Classification.** Many systems of classification have been proposed for temporomandibular disorders. The most recent classification was offered by the American Academy of Craniomandibular Disorders in 1990 (1). This system is particularly useful for the medical community because it incorporates specific pathological states of the joint in the International Headache Society's Classification and Diagnostic Criteria for Headache Disorders, Cranial Neuralgias and Facial Pain. It is also useful for the dental profession because it provides the practitioner with a realistic perspective on how these disorders fit among the many causes of head and neck pain.

The disorders can be separated into two general categories, those affecting the associated musculature and those primarily affecting the joint (internal derangement). If one looks at the various subclassifications of these two groups of disorders (Table), their similarity to disorders affecting other joints and muscles throughout the body becomes evident. What differentiates these disorders from those of other joints is that the temporomandibular joints function in tandem; both left and right condyles are connected by a single bone. Furthermore, the function of the joints can be affected by the position of the jaw, which is in part dictated by the contact of the teeth (occlusion). Investigation of such problems requires thorough evaluation of the patient by a dental practitioner with expertise in these disorders.

**Evaluation.** Patients suspected of having one of these disorders need to undergo a complete evaluation. This involves both medical and dental histories; examination of the head and neck region, including palpation of the muscles of mastication and temporomandibular joints; evaluation of mandibular range of motion; and auscultation of joint sounds (2). Dentition must also be examined for the presence of carious lesions, periapical pathology, and periodontal disease. This allows the practitioner to rule out den-

---

From the Office of the Director of the Temporomandibular Disorder Clinic (AK) and the Division of Oral and Maxillofacial Surgery (DB, SM), The Mount Sinai Hospital, New York, NY. Address reprint requests to Andrew Kaplan, DMD, Box 1187, One Gustave Levy Pl., New York, NY 10029.

TABLE

*Comparison of Diagnostic Classification of Muscular and Temporomandibular Disorders*

| Muscular Disorders | Temporomandibular Disorders |
|---|---|
| **Myofascial pain** | **Deviation in form** |
| **Myositis** | **Disc displacement**<br>With reduction<br>Without reduction |
| **Sprain** | |
| **Reflex splinting** | **Hypermobility** |
| **Contracture** | **Dislocation** |
| **Hypertrophy** | **Inflammatory conditions**<br>Synovitis<br>Capsulitis |
| **Neoplasm** | |
| | **Arthritides**<br>Osteoarthrosis<br>Polyarthritides |
| | **Ankylosis**<br>Fibrous<br>Bony |

toalveolar abnormality as the cause of the facial pain. Evaluation of the cervical range of motion is indicated because cervical pain often accompanies temporomandibular joint pain (3).

Imaging of the joints may be necessary. The particular technique used will depend on whether osseous or soft-tissue abnormality is suspected. Panoramic radiography and lateral oblique plain film views are the most basic techniques of evaluating osseous structures (4). Computerized tomography can also be used for this purpose. Arthrography and magnetic resonance imaging are the most useful techniques for evaluating the joint's soft tissues (5). In rare instances, technetium bone scanning may be necessary to determine the activity of an arthritic or neoplastic process.

Blood and urine analyses are occasionally needed, particularly if there is suspicion of systemic disease, such as rheumatoid arthritis, psoriatic arthritis, ankylosing spondylitis, gouty arthritis, or chondrocalcinosis.

There has recently been an explosion of sophisticated electronic diagnostic equipment, including electromyographic devices, thermographic units and jaw-tracking equipment, on the market for diagnosis of temporomandibular disorders. These instruments are not necessary for proper diagnosis or treatment, and the meaning of the data generated by such instrumentation has yet to be defined through well-designed scientific study (6).

## Muscular Disorders of the Temporomandibular Joint

**Trigger Points.** Muscles throughout the body can develop discrete points or bands, known as trigger points, that cause severe pain when stimulated. Trigger points can refer pain to seemingly unrelated areas. For the muscles of mastication, pain referral areas may include the teeth, sinuses, ears, and throat. These trigger points tend to be self-perpetuating (7), leading to a chronic state of pain.

Although pain referral from trigger points probably does not occur as frequently as was once believed, it is important to recognize the possibility of such an event. Many patients have undergone unnecessary root canal therapy or tooth extraction as a result of misdiagnosed trigger points.

**Myofascial Pain Dysfunction Syndrome.** Myofascial pain dysfunction syndrome is among the most common disorders affecting the muscles of mastication (8). This dysfunction is manifested as masticatory muscle soreness resulting from overuse, usually from clenching or grinding (bruxism) of the teeth.

Patients with this condition often complain of facial pain or soreness on awakening; limited and painful movement of the jaw; a fatigued feeling when eating; and soreness or headache in the temporal region. The pain is often provoked by intentional contraction of the suspected muscle. For example, pain on protruding the mandible isometrically may indicate lateral pterygoid involvement. Specific muscle soreness is indicative of the patient's particular pattern of grinding or clenching (parafunctional habits).

The diagnostic criteria for the syndrome include at least three of the following: facial pain, usually but not always unilateral; pain on palpation of one or more muscles of mastication; limited or deviated movement of the mandible; and absence of radiographic changes in the joints themselves (9).

The most common treatment for the syndrome includes the occlusal splint, a plastic device custom-fitted over the upper or lower teeth and intended to prevent the teeth from coming together (occluding). These splints are thought to "deprogram" the muscles of mastication and have been shown to be approximately 80% effective in reducing muscle activity.

If stress plays a major role in the disorder, consultation with a psychiatrist or psychologist may be indicated. Biofeedback or relaxation training is most beneficial. Physical therapy is particularly helpful for patients who experience significant cervical pain or for patients who suffer from poor postural habits.

The use of benzodiazepines to help control bruxism or clenching should be reserved for short-term therapy (two weeks maximum), in order to avoid unnecessary dependence. Nonsteroidal antiinflammatory agents are also useful in patients who have gone on to develop myositis. In patients who are not successfully treated by these techniques, antidepressant medications may be added to the regimen. Treatment often requires a multidisciplinary approach.

**Myositis.** Inflammation of the muscles of mastication often accompanies myofascial pain dysfunction syndrome. Myositis can be a primary finding, particularly after acute traumatic injury to one of the muscles of mastication (10), or it may become a secondary finding in patients with long-standing myofascial pain dysfunction. Some patients may develop clinically observable swelling of the muscle. These muscles may be involved in whole or in part.

**Masseteric Hypertrophy.** Patients who have long-term parafunctional habits can develop unilateral or bilateral enlargement of one or both masseter muscles (11). Many of these patients are asymptomatic but display obvious facial asymmetry which is observed on routine examination or brought to a patient's attention by a friend or family member. After the physician has ruled out neoplastic disease by history and physical examination, and occasionally by imaging with computerized tomography (CT) or magnetic resonance imaging (MRI), these patients need only reassurance that the process is benign. If a significant esthetic problem is the chief complaint, surgical reduction of the masseter muscles may be a treatment alternative. If the parafunctional habit continues, the problem may recur.

**Contracture.** As in other parts of the body, temporomandibular muscles subject to disuse may develop contractures. The short-term change in the muscle is a decrease in fiber length. This reversible process is known as myostatic contracture, and occurs as a painless hypomobility and deviation to the affected side on opening. The condition is commonly seen following removal of maxillomandibular fixation, such as after a six-week healing period following mandibular fracture. Active and passive stretching can remedy myostatic contracture.

With long-term disuse, muscle fibers can be replaced by fibrous tissue. This process is irreversible. Stretching exercises will thus result in tearing of the muscle fibers. If the limited mandibular movement is severe and the patient is unable to carry on normal activities, surgical release may be necessary.

**Muscle Guarding.** Muscle contracture can also be a result of "muscle guarding," a voluntary limitation of mandibular use. A common example is the patient with temporomandibular joint osteoarthrosis. Patients voluntarily limit the use of the mandible because they experience pain throughout active range of motion. In this situation, the primary cause of muscle guarding must be remedied before full range of motion can be restored.

**Neoplastic Disease.** The muscles of mastication are subject to benign and malignant changes (for example, myxoma, rhabdomyosarcoma). When the patient's chief complaint is swelling, neoplastic disease of the salivary glands (12) should be suspected. Central nervous system tumors may impinge on the trigeminal nucleus, producing changes in masticatory muscle function.

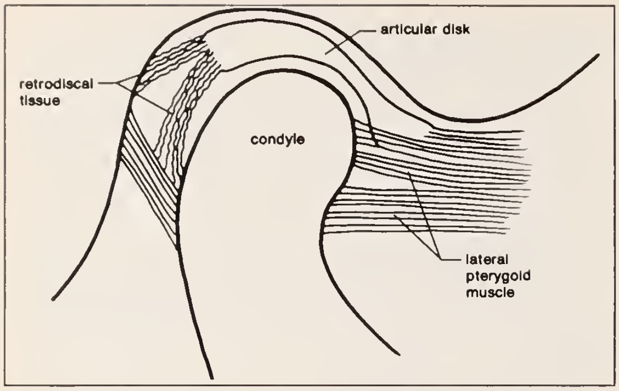
## Internal Derangements of The Joint

The temporomandibular joint is made up of three major components, the condyle, the temporal component or fossa, and the meniscus or disc. The joint is also surrounded by a fibrous capsule and is limited in its normal movements by several ligaments. Internal derangements of the joint are a group of disorders affecting the structure and function of the joint. These include changes in the meniscus, ligaments, and osseous structures.

**Capsulitis.** Inflammation of the fibrous capsule is called capsulitis and is often caused by trauma but can also be a secondary finding in patients with internal derangement or osteoarthrosis. Inflammation of the synovium is called synovitis. Since treatment is the same for both entities, differentiation between these two conditions is academic.

Patients with capsulitis have pain on palpation of the lateral aspect of the joint as well as pain on active and passive movement. They also may have muscle guarding. Radiographically, the joint appears normal. Capsulitis is best treated by placing the patient on a soft diet and nonsteroidal antiinflammatory medications. Voluntarily limiting opening of the mouth is encouraged. If the pain has not subsided in two or three weeks, occlusal splint therapy may be necessary.

**Disc Displacement.** Disc displacement is probably the most common form of internal derangement. In a normal joint, the disc is divided into three portions: a posterior thick band, a thin



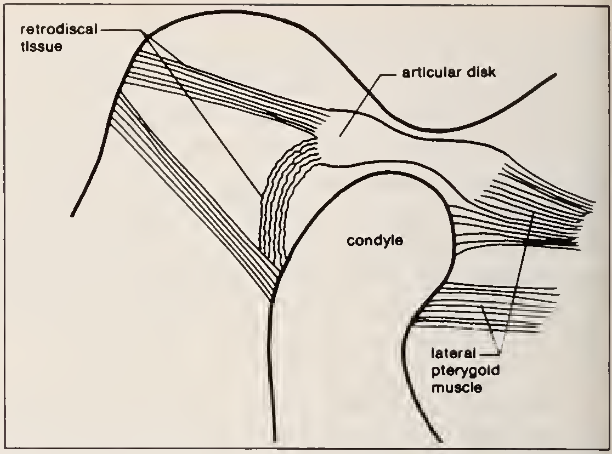
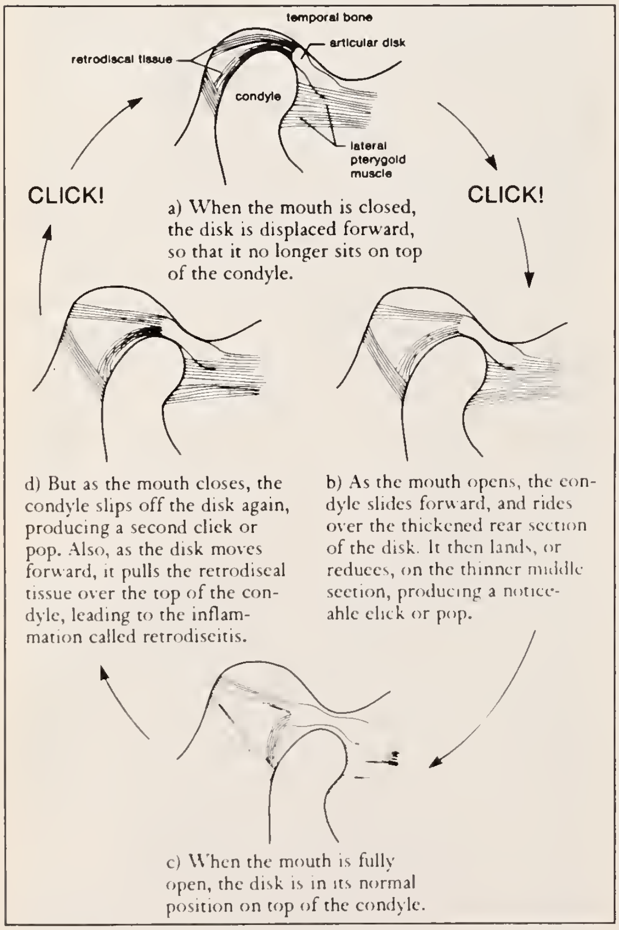


**Fig. 1. Left** Normal temporomandibular joint with mouth in closed position, sagittal view. **Fig. 2. Right** Normal temporomandibular joint with mouth in full open position.

intermediate band, and an anterior band of intermediate thickness. When the mouth is closed (Fig. 1), the thick band band sits on the head of the condyle in the twelve o'clock position. When the mouth is opened, the condyle translates and rotates downward and forward as the disc rotates with it. On maximal opening of the mouth, the condyle is positioned opposite the intermediate thin band (Fig. 2).



a) When the mouth is closed, the disk is displaced forward, so that it no longer sits on top of the condyle.

b) As the mouth opens, the condyle slides forward, and rides over the thickened rear section of the disk. It then lands, or reduces, on the thinner middle section, producing a noticeable click or pop.

c) When the mouth is fully open, the disk is in its normal position on top of the condyle.

d) But as the mouth closes, the condyle slips off the disk again, producing a second click or pop. Also, as the disk moves forward, it pulls the retrodiscal tissue over the top of the condyle, leading to the inflammation called retrodiscitis.

Fig. 3. Internal derangement with reduction.

When the disc is displaced, it usually moves in an anterior and medial direction (although lateral and posterior displacements have been reported). On opening, the condyle "jumps" over the anteriorly displaced thick band, giving rise to the characteristic, commonly heard click (Fig. 3). On closing, the condyle slips off the disc, causing a second or reciprocating click. This is known as anterior disc displacement with reduction (13). As the ligaments that hold the components of the joint in close apposition become stretched, they allow for further forward displacement of the disc. As this happens, the click will occur at a later and later phase of opening. Thus, the severity of the displacement can be estimated clinically by the time in which the click occurs during opening. Some clicking joints, particularly those not accompanied by pain or limited movement, need no treatment.

At some point the disc may be so far displaced that the patient may begin to complain of periodic locking or an inability to open the mouth fully. This is called anterior disc displacement with intermittent locking (14). These periods may begin to increase in frequency and length until the patient's mouth is locked all the time, now called anterior disc displacement without reduction. The limitation of movement is due to displacement of the disc in such a way that it obstructs condylar translation (Fig. 4).

Disc displacements accompanied by pain are probably due to inflammation behind the posterior band of the disc, known as the retrodiscal tissue. Anterior displacement of the disc gives rise to inflammation and pain as the highly vas-

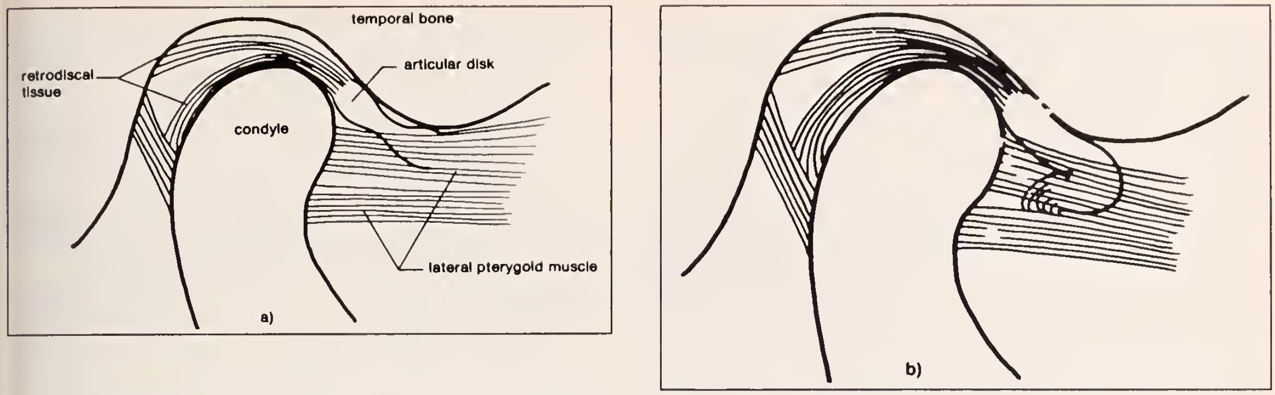


Fig. 4. Internal derangement without reduction: (a) mouth in closed position; (b) mouth in open position.

cular and well innervated retrodiscal tissue moves forward.

Disc displacement can be treated in a variety of ways. Generally, nonsurgical treatment should be tried before resorting to surgical care. Nonsurgical treatment approaches are similar to those used for the muscular disorders previously described. Some authorities advocate the use of "repositioning splints" or devices aimed at changing the spatial relationship of the disc, condyle, and fossa in an attempt to recapture the disc. These techniques can be involved, because they often require extensive orthodontic or prosthodontic therapy at some later date. Although these techniques are beneficial for some patients, their use on all patients with disc displacements is not warranted.

Some authorities advocate the use of intracapsular steroid injections. Although a single injection may be beneficial, serial injections are not advocated because joint degeneration may be accelerated (15).

Surgical approaches are common therapeutic technique for disc displacement. Arthrotomy and in particular disc plication, in which the disc is repositioned, can be effective. Since the mid-eighties, arthroscopic procedures have been popular and effective because of their relative simplicity and minimal associated morbidity. Arthroscopic surgery has been particularly useful for patients with anterior disc displacement without reduction and other internal derangements. It allows for increased mobility and reduction of pain.

The ability of an arthroscopic procedure to actually reposition a displaced disc is doubtful. On the other hand, proponents of arthroscopic surgery question the need to reposition the disc for successful resolution of pain and improvement in function.

Disc plication surgery or other "open joint surgery" has been largely relegated to those patients for whom arthroscopic procedures have failed, those who have a poor prognosis, or those who have more complex problems than can now be improved with arthroscopy.

**Retrodiscitis.** The highly vascular retrodiscal tissue is subject to severe inflammation when blunt trauma is sustained, forcing the condyle superiorly and posteriorly into this area, a condition called retrodiscitis. The problem is generally acute unless accompanied by disc displacement or present in a severe bruxer who is constantly loading the tissues. Short-term splint wear, nonsteroidal antiinflammatory therapy, soft diet, and voluntary limitation of movement are generally the treatments of choice.

**Arthritides.** Most forms of arthritis that affect other joints in the body can affect the temporomandibular joint. Although space does not allow a full review of the arthritides, some of the salient features are as follows.

The rheumatoid arthritic patient can be subject to severe degeneration of the condyles with subsequent ramus shortening, resulting in an anterior open bite. The patient will note shifting in the occlusion and difficulty taking complete bites of food. The end stage may result in fibrous or bony ankylosis of the joint and related hypomobility (16). Juvenile rheumatoid arthritis can affect the temporomandibular joint by slowing or halting growth of the mandible, evidenced by severe retrognathia. Involvement of the temporomandibular joint has been reported in psoriatic arthritis, ankylosing spondylitis, gouty arthritis, and chondrocalcinosis. In addition, the polyarthritis of Lyme disease can also affect the temporomandibular joint (17).

By far the most common arthropathy to affect this joint is osteoarthrosis or degenerative joint disease. The associated bony changes are similar to those found throughout the body. The osseous alterations are thought to occur in response to demands on the joint exceeding its capacity to adapt. Etiologies include the microtrauma of long-term clenching, bruxing, and other events. Preexisting disc displacement has also been implicated (18). The disease generally progresses through an acute stage during which the osseous changes are taking place. Pain throughout the mandibular range of motion may be present. This phase can last from six to eighteen months. The acute phase is painful and should be managed symptomatically. The second or chronic phase of the disease is usually characterized by painless adequate function that may be accompanied by crepitus (19).

**Developmental Disturbances.** The temporomandibular joints are subject to several developmental disturbances, including failure of one or both condyles to develop (condylar aplasia) and incomplete development of the condyle (condylar hypoplasia). If either problem is unilateral, the patient will have an evident facial asymmetry that becomes apparent in childhood. Functionally, the mandible will deviate toward the affected side on opening (20).

Overdevelopment of the condyle, or condylar hyperplasia, can be either congenital or acquired. The patient will have an evident facial asymmetry but on opening will deviate away from the affected side. When this disorder is acquired over a short period, the patient will complain of bite shifting and have a notable separation of the posterior teeth on the affected side.

Developmental defects of the face often involve several bones and can include the condyle, temporal bone, and maxilla, illustrating the interdependence of these bones during development. If several bones are involved, the developmental problem may be labeled hemifacial microsomia (21).

**Neoplastic Disease.** Like the muscles of mastication, the temporomandibular joints are subject to neoplastic disease. Fortunately malignancy is very rare. Common benign tumors affecting the joint include osteoma, chondroma, osteochondroma, and synovial chondromatosis (22).

### References

1. McNeil C. Craniomandibular disorders, guidelines for evaluation, diagnosis and management. Chicago, Ill: Qintessence, 1990, 13–19.
2. Kaplan A. Examination of the TMJ patient. In: Kaplan A, Assael L, eds. Textbook of temporomandibular disorders. Philadelphia: W B Saunders (in press).
3. Kraus SL. Cervical spine influence on the craniomandibular region. In: Kraus SL, ed. TMJ disorders: management of the craniomandibular complex. New York: Churchill Livingstone, 1988, 367–404.
4. Kaplan A. Plane radiography of the TMJ. In: Kaplan A, Assael L, eds. Textbook of temporomandibular disorders. Philadelphia: W B Saunders (in press).
5. Milbauer D. CT and MRI imaging of the TMJ. In: Kaplan A, Assael L, eds. Textbook of temporomandibular disorders. Philadelphia: W B Saunders (in press).
6. Mohl N, McCall W, Lund J, Plesh O. Devices for the diagnosis of temporomandibular disorders. Part I: Introduction, scientific evidence and jaw tracking. J Pros Dent 1990; 63(2):198–201.
7. Travel J, Simons D. Myofascial pain and dysfunction: the trigger point manual, chap. 2. Baltimore: Williams and Wilkins, 1983, 5–44.
8. Fricton J, Kroening R, Haley D. Muscular disorders, the most common diagnosis. In: Fricton J, Lroenig R, Hathaway K, eds. TMJ and craniofacial pain, diagnosis and management, chap. 7. St Louis: Ishayaku EuroAmerica, 1988, 67–83.
9. Laskin D. Etiology of the pain dysfunction syndrome. JADA 1969; 79:147–153.
10. Bell W. Temporomandibular disorders, 2nd ed. Chicago: Year Book Medical Publishers, 1986, 181.
11. Shaefer W, Hein H, Levy B. A textbook of oral pathology, 4th ed. Philadelphia: W B Saunders, 1983, 875.
12. Kaplan A, Som P, Lawson W. The coexistence of myofascial pain dysfunction syndrome and mucoepidermoid carcinoma of the parotid gland. JADA 1986;112:495–496.
13. Okesson J. Management of TM disorders and occlusion 2nd ed. St. Louis: C V Mosby, 1989, 285.
14. Fricton J, Kroening R, Haley D. Joint disorders: derangement and degeneration. In: Fricton J, Lroenig R, Hathaway K, eds. TMJ and craniofacial pain, diagnosis and management. St. Louis: Ishayaku EuroAmerica, 1988 85–130.
15. Poswillo D. Experimental investigation of the effects o intraarticular hydrocortisone and high condylectom on the mandibular condyle. Oral Surg Oral Med Ora Path 1970; 30:161–165.
16. Etalla-Ylitalo UM, Syrjanen S, Halonen P. Functiona disturbances of the masticatory system related to temporomandibular joint involvement by rheumatoid arthritis. J Oral Rehab 1987; 14:415–420.
17. Harris R. Lyme disease involving the temporomandibula joint. J Oral Maxillofac Surg 1988; 46:78–82.
18. Kaplan A, Buchbinder D. Arthritis of the TMJ. In: Kapla A, Assael L, eds. Textbook of temporomandibular disorders. Philadelphia: W B Saunders, (in press).
19. Mejersjo C. Therapeutic and prognostic considerations i TMJ osteoarthrosis: a literature review and a long ter study in 11 subjects. J Cranio Practice 1987; 5(1):70–7
20. Shaefer W, Hein H, Levy B. A textbook of oral patholog 4th ed. Philadelphia: W B Saunders, 1983, 702–713.
21. Assael L. Developmental disturbances of the TMJ. I Kaplan A, Assael L, eds. Textbook of temporomandibular disorders. Philadelphia: W B Saunders, (in press
22. Shaefer W, Hein H, Levy B. A textboook of oral patholog 4th ed. Philadelphia: W B Saunders, 1983, 712.

# Diagnosis and Treatment of Myofascial Pain

HANS KRAUS, M.D., AND ANDREW A. FISCHER, M.D., PH.D.

MANY pain syndromes elude diagnosis because the results of fundamental and sophisticated tests, such as magnetic resonance imaging and electromyograms, are usually normal, and clinical examination frequently neglects the skeletal muscle. Mysterious back and neck problems, often misdiagnosed as mechanical in origin, as well as injuries to the extremities, are found to have a muscular, not an orthopedic, cause.

This article reviews the diagnosis and treatment of painful conditions that originate in muscles. Quantitative evaluation of tenderness by pressure algometry and measurement of muscle spasm by tissue compliance meter are described. These parameters, each measurable by an instrument developed by one of the authors, significantly advance the diagnosis of pain.

## Types of Muscle Pain and Their Treatment

To correctly assess and treat muscle-related problems, the four types of muscle pain—tension, spasm, muscle deficiency (weakness), and trigger points—must be distinguished.

**Tension.** Tension is characterized by excessive, painful, sustained muscle contraction (1).

The most effective treatment is relaxation exercise (2), with tranquilizers as adjunct therapy when indicated. However, to control tension and prevent recurrence, one must root out the cause, whether emotional or extrinsic.

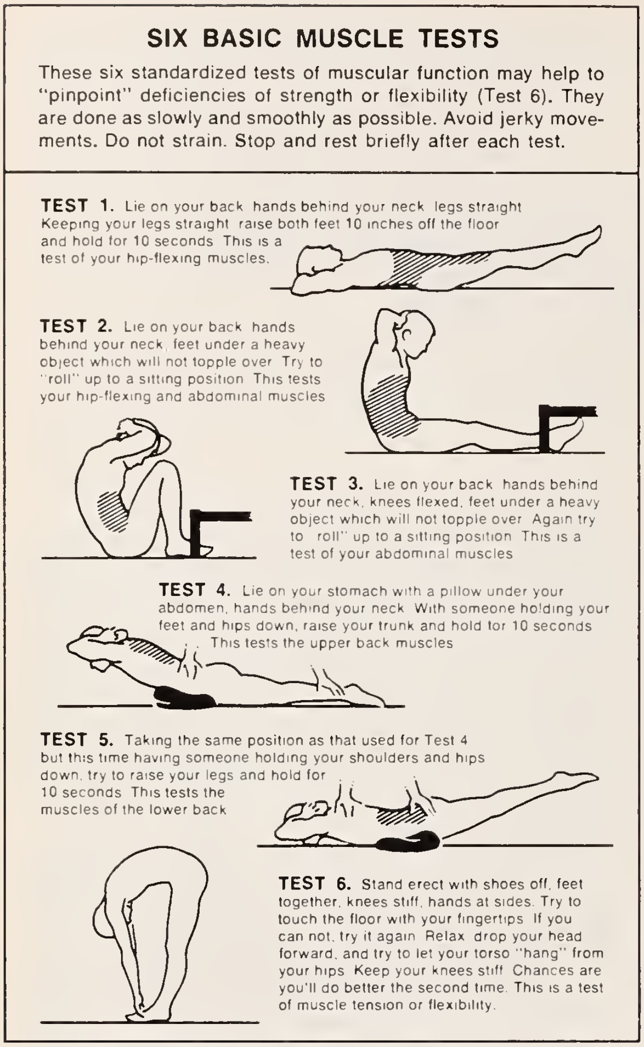
**Muscle Spasm.** Muscle spasm has been defined as a sustained, involuntary contraction that is usually intrinsically painful and cannot be relieved by effort (3). The most frequent causes of spasm are direct trauma, sprains, strains, fractures, and radiculopathy. Unlike the muscle contraction caused by tension, muscle spasm inhibits motion such that every attempt to move increases pain. Spasm responds poorly to tranquilizers and analgesics.

Treatment consists of gentle, controlled movement made possible by a vapocoolant spray such as ethyl chloride (4). More effective than Fluorimethane (dichlorodifluoromethane, 15%, and trichloromonofluoromethane, 85%, Gebauer Chemical Co., Cleveland, OH) (5), ethyl chloride is sprayed directly onto the most painful area. Gentle limbering motion is then initiated. Because pain usually shifts during the spray and limber procedure, spraying the new area of discomfort makes further movement possible. Treatment is continued until no further increase in range of motion or relief of pain can be obtained. This procedure, known as "spray and limber," involves the patient's active movements (6), as opposed to "spray and stretch" (5, 10), which elongates the involved muscle passively.

Treatment of back pain utilizes the prone position (6). One pillow is placed under the patient's hips and another under the legs. The painful area is sprayed with ethyl chloride as the patient's buttocks are tightened and then relaxed; this simultaneously contracts and relaxes the sacrospinous, gluteus, and quadratus lumborum muscles. When the pain subsides, the patient is allowed to assume the lateral decubitus position. The head and knees are supported with pillows. The patient is instructed to slowly flex and extend the hip and knee while coolant spray is applied to the painful area. Four or five repetitions of the exercise are

---

From the Department of Rehabilitation Medicine, Mount Sinai School of Medicine (CUNY), New York, NY and the Office of the Chief, Rehabilitation Medicine Service, Veterans Administration Medical Center, Bronx, NY. Address all correspondence and reprint requests to Andrew A. Fischer, Chief, Rehabilitation Medicine Service, Veterans Administration Medical Center, 130 W. Kingsbridge Road, Bronx, NY 10468.



**SIX BASIC MUSCLE TESTS**

These six standardized tests of muscular function may help to "pinpoint" deficiencies of strength or flexibility (Test 6). They are done as slowly and smoothly as possible. Avoid jerky movements. Do not strain. Stop and rest briefly after each test.

**TEST 1.** Lie on your back hands behind your neck legs straight Keeping your legs straight raise both feet 10 inches off the floor and hold for 10 seconds This is a test of your hip-flexing muscles.

**TEST 2.** Lie on your back hands behind your neck, feet under a heavy object which will not topple over Try to "roll" up to a sitting position This tests your hip-flexing and abdominal muscles

**TEST 3.** Lie on your back hands behind your neck, knees flexed, feet under a heavy object which will not topple over Again try to roll" up to a sitting position This is a test of your abdominal muscles

**TEST 4.** Lie on your stomach with a pillow under your abdomen, hands behind your neck With someone holding your feet and hips down, raise your trunk and hold for 10 seconds This tests the upper back muscles

**TEST 5.** Taking the same position as that used for Test 4 but this time having someone holding your shoulders and hips down, try to raise your legs and hold for 10 seconds This tests the muscles of the lower back

**TEST 6.** Stand erect with shoes off, feet together, knees stiff, hands at sides. Try to touch the floor with your fingertips If you can not, try it again Relax drop your head forward, and try to let your torso "hang" from your hips Keep your knees stiff Chances are you'll do better the second time. This is a test of muscle tension or flexibility.

**Fig. 1.** Kraus-Weber test for strength and flexibility of key postural muscles.

generally required. The patient then turns onto the other side and the procedure is repeated.

This exercise and others should be performed hourly at home. Patients are encouraged to avoid prolonged sitting, standing, or walking. In contradistinction to management of other types of pain, return to work is advisable in most cases. Hospitalization and prolonged bed rest are counterproductive to those with muscle spasm pain.

Electrical tetanizing-current stimulation to the spasmodic muscle for 10 min interrupts the spasm by fatiguing the muscle. The tetanizing current is followed by 10 min of sinusoid surging current, which induces periodic contractions, increases blood flow, and relaxes muscles. The electrical stimulation is followed by spray and limbering exercises as described above.

This protocol of electrical stimulation and limbering exercise is then incorporated into a physical therapy program. Therapy is provided daily in the acute stage and three times weekly until functional recovery and pain relief are achieved.

**Muscle Deficiency.** Muscle deficiency is characterized by weakness and stiffness (loss of flexibility), stemming from inactivity due to injury, immobilization, or sedentary living. It responds favorably to exercise.

For back pain, exercises are prescibed to remedy deficiencies identified in results of the Kraus-Weber test (7), which measures strength and flexibility of key postural muscles (Fig. 1). Weakness and stiffness of abdominal, back, and hamstring muscles are the most frequently encountered problems. Abdominal weakness appears to be more prevalent than back weakness.

A customized exercise program must take into accunt the tension that accompanies muscle deficiency. Exercises should follow a pattern of relaxation, warmup, strengthening, and stretching. Usually no exercise is repeated more than three or four times in succession. New exercises are added gradually in the sequence mentioned above, and the exercise order is reversed as a cooloff.

**Trigger Points.** First described comprehensively by Lange (8) in 1931, trigger points have gradually been accepted as significant over the past few decades. Since the 1940s, descriptions have appeared with increasing frequency.

These points are small, exquisitely tender areas that spontaneously or on activation may shoot pain into a remote area, the "reference pain zone" (5, 9, 10). Each trigger point has a specific reference zone; for example, if the point is in the gluteus minimus, it may shoot pain down over the lateral and posterior lateral aspect of the lower extremity. Such pain patterns frequently lead to the erroneous diagnosis of sciatica (lumbosacral radiculopathy) when the pain is actually caused by muscle trigger points that are easily treated.

*Diagnosis* of trigger points relies on abnormal tenderness, which can be quantified by a pressure algometer (dolorimeter) (11). A pathognomonic finding for myofascial trigger points is palpation of a "taut band," which consists of hypersensitive muscle fibers of a consistency harder than normal (5, 10). Fibromyalgia, previously called fibromyositis, is also characterized by tender points, though these differ from myofascial trigger points in several respects. Fibromyalgic tender points are widespread and symmetric; myofascial trigger points are confined to one region (12–17) and usually affect one side only (5, 6, 10, 15–17). Fibromyalgic tender points have characteristic locations, including tissues other than muscles (12–14); they do not produce referred pain, and taut bands are absent (10, 14, 15).

Trigger points are located most frequently in muscle or fascia and are therefore called myofascial. Nonetheless, point tenderness can be detected in sprained ligaments, tendons, pericapsular tissue, or even the periostium.

Typical myofascial trigger point locations include (6, 10) suboccipital muscles (causing occipital headache); neck muscles (producing pain that radiates to the arm and forearm); infraspinatus muscles (causing pain radiating to the posterior arm); and rhomboid muscles (causing interscapular pain). Trigger points often result from innocuous acts such as squeezing a telephone between the ear and shoulder. They may also contribute to temporomandibular joint problems.

Trigger points in sacrospinous and quadratus lumborum muscles often combine with those in glutei muscles and tensor facia lata to produce pain that radiates to the lower extremities. This complex scenario may cause peripheral nerve entrapment, which closely resembles disc disease. Myotomal distribution of trigger points indicates irritation (compression, radiculitis) of the supplying nerve root. Thus this condition is frequently diagnosed erroneously as discogenic pain (the pseudodisc syndrome) (6).

A number of factors can produce trigger points, the most important being acute or chronic local trauma; sprain, such as occurs in sport injuries; motor vehicle accidents; and whiplash injuries.

Trauma-induced trigger points are frequent. When localized at the lateral or medial humeral condyle, they are diagnosed as "tennis" or "golf" elbow. After knee injury, they arise at the vastus medialis or gastrocnemii. After ankle injury, they appear in the soleus and gastrocnemius muscles.

*Treatment:* The most effective treatment of trigger points consists of an injection that combines infiltration using a local anesthetic with "needling," the repetitive insertion and withdrawal of a needle. The entire taut band, which frequently extends from the origin to the insertion of the muscle, should be injected and needled for best results. This procedure mechanically breaks up the entire inflamed or fibrotic tissue, inducing a long-term cure. The anesthetic effects of 1% lidocaine last approximately 1 hr. Clinical experience of many years shows that once trigger points become chronic and damaged tissue has healed by fibrotic scarring, mechanical breakup of taut bands by needling is necessary for complete and long-lasting healing. A successful injection stops the pain immediately. Alleviation of pain following injection and needling verifies that the complaint had been caused by the injected trigger point. Each point requires at least one injection. Consequently, patients with pain in several involved areas are treated by a series of injections given once or twice weekly.

Each injection is followed up by three sessions of specific physiotherapy, which enhances and accelerates the healing process. Twenty-minute hot packs relax the injected muscle(s). Electrical stimulation by sinusoid surging currents (for 15 min) induces strong contraction of the muscle, thus squeezing out the edema caused by needling and increasing blood flow. The augmented blood flow purges products of tissue injury, thereby preventing inflammation. Ethyl chloride spray to painful points inactivates the remaining trigger points and is followed by limbering exercises to restore range of motion.

Physical therapy is most effective when performed on the days immediately following the injections. In addition, the patient in the acute stage performs limbering exercises and stretching of the injected muscle every hour. Later, limbering is repeated less frequently. Spraying with vapocoolant inactivates trigger points and, if followed by limbering exercises (2, 6) or stretching (10), is also effective when injection is not feasible.

*Treatment of factors causing trigger points and prevention of recurrences:* Unless the factors causing or perpetuating trigger points are removed, recurrence is likely (6, 10). In a patient with muscle spasm, trigger point formation can be prevented by immediate controlled mobilization rather than by immobilizing the painful area (4, 6).

Causes of myofascial trigger points as well as perpetuating factors can be mechanical or systemic (10). Mechanical causes and perpetuating factors include leg length discrepancy (which should be corrected by a lift); postural stress in sitting or standing; and vocational stress in overloading specific muscles and repetitive movements.

Systemic factors may induce irritability of the skeletal muscles and inability to relax. Such conditions include hypothyroidism (6, 9, 10); estrogen deficiency (6, 9); metabolic dysfunction; chronic viral, bacterial, or parasitic infection; enzyme dysfunction; electrolyte disturbance; vitamin deficiency; and psychological stress (10).

## Quantitative Diagnosis

**Algometry.** A major advance in diagnosing and treating myofascial pain is the capability of objectively quantifying tenderness, trigger points (11, 18–21), and fibromyalgia (12, 13). Pocket-sized pressure algometers are used for this pur-

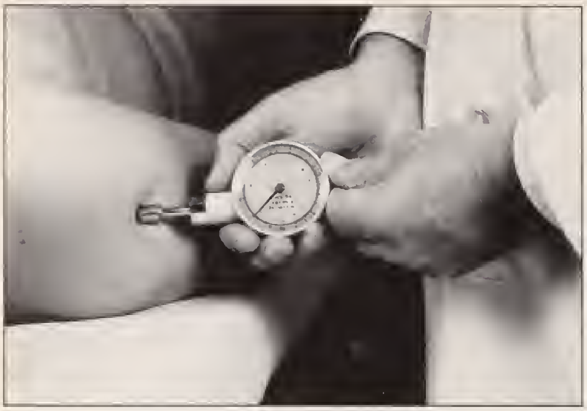


**Fig. 2.** Pressure threshold meter, an algometer proven effective in quantitative diagnosis of abnormal tenderness.

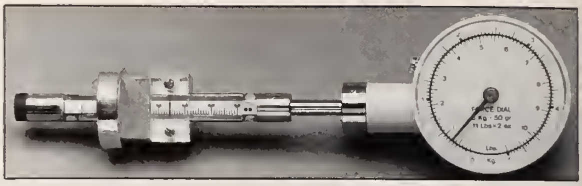


**Fig. 4.** Tissue compliance meter for objective quantitative measurement of changes in muscle tone (consistency).

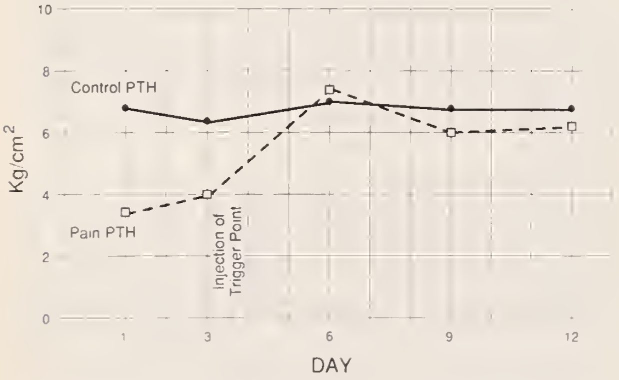
pose. They are inexpensive, simple to use, and provide valuable information that cannot be obtained by other methods. Muscle consistency diagnostic of spasm and trigger points can also be assessed quantitatively and objectively by tissue compliance meters.

One type of algometer is the pressure threshold meter (11) (Fig. 2). It is a small mechanical force gauge, fitted with a rubber tip measuring 1 $cm^2$ exactly. First, the maximum tender point is located through careful palpation and is marked. Then the meter tip is applied to the spot. The patient is instructed to indicate when pain is first perceived. Pressure is then increased at the rate of 1 kg/sec until pain or discomfort occurs; the minimum force that causes pain is termed the pressure pain threshold. A threshold reading of 2 kg less than the opposite (normal) area or an adjacent nontender spot indicates abnormal pressure sensitivity (18, 19). Patients sometimes feel discomfort in a “reference pain zone” (10) remote from the underlying trigger points; the trigger point causing the referred pain must also be identified and treated.

Normal values for pressure threshold over the muscles most frequently afflicted by trigger points have been established (20). The reliability, validity, and reproducibility of the meter readings for diagnosis of tender spots has been proven (18, 19, 21, 22). The meter has been used successfully for evaluation of treatment results after spray and stretch therapy (23), transcutaneous electrical nerve stimulation (24), and physical therapy (25). Hypersensitive points can often be documented objectively as hot spots on thermogram (heat imaging) (26).

Pressure threshold measurements have two clinical uses (15):

- **Quantitative diagnosis of tenderness**
  - tender spots: trigger points, fibromyalgia
  - activity of inflammation in arthritis
  - diagnosis of low pain tolerance
- **Measurement of treatment results**
  - immediately after procedure
    - trigger point injections
    - nerve blocks
  - long-term effects



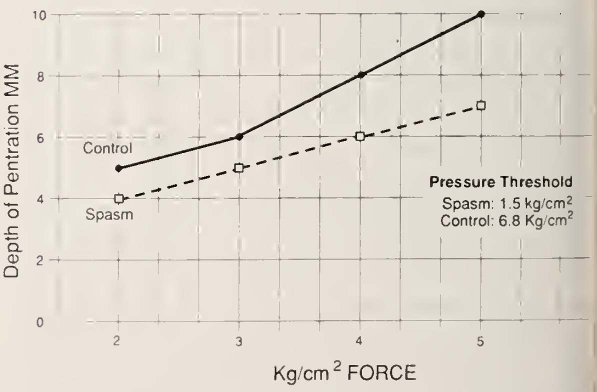


**Fig. 3.** Left Changes in pressure threshold after injection at trigger point in upper trapezius muscle of 39-year-old man wit neck pain for 5 mo caused by cervical radiculopathy. Pressure threshold: day 1, painful side (Pain PTH), 3.4 kg/cm$^2$; oppposit normal side (Control PTH), 6.8 kg/cm$^2$; day 3, prior to injection, painful side, 4 kg/cm$^2$; opposite side, 6.4 kg/cm$^2$. Day 3, injection 1% lidocaine (5 mL) and needling. Pressure threshold: day 6, painful side, 7.4 kg/cm$^2$; control side, 7.0 kg/cm$^2$. Followup, day 9 injected point, 6.0 kg/cm$^2$; control side, 6.8 kg/cm$^2$; day 12, injected point, 6.2 kg/cm$^2$; control side, 6.8 kg/cm$^2$. **Fig. 5.** Righ Tissue compliance measurements over right upper trapezius muscle in spasm in 44-year-old man with upper back pain cause by cervical radiculopathy and root irritation. Palpation revealed contracted and tender right upper trapezius muscle. Pressur threshold measured at most sensitive point of spasmodic muscle (Spasm) was significantly lower, indicating harder consistenc than at opposite (Control) side. Hypersensitive site indicates presence of trigger point.

General pain sensitivity can be diagnosed by algometry, and the relation of that sensitivity to focal tenderness can be established (27). Abnormal pressure pain sensitivity can be differentiated (Fig. 3); moreover, short-term and long-term effects of treatment can be documented quantitatively (Fig. 3).

**Tissue Compliance Recording.** The tissue compliance meter (Fig. 4) is a hand-held mechanical instrument that objectively quantifies muscle consistency (28). The meter consists of a force gauge featuring a long shaft ending in a rubber disk (1 $cm^2$), which is pressed into the examined tissue. A sliding disk fitted around the shaft of the force gauge indicates the depth of penetration. Compliance is expressed as the depth the rubber disk penetrates at a known force. This meter can record the harder consistency of a muscle in spasm without invading the skin. It can also quantitatively document relief of spasm after nerve block (28, 29).

Tissue compliance measurement can objectively document muscle spasm and the presence of a trigger point (Fig. 5).

## Summary

The importance of muscle as a source of pain should not be underestimated. Proper diagnosis of the four types of muscle pain (tension, spasm, muscle deficiency, and trigger points) is essential to effective management. Pain-diagnostic instruments can quantify tenderness and spasm and help evaluate treatment results.

### References

1. Sainsbury P, Gibson JG. Symptoms of anxiety and tension and accompanying physiological changes in the muscular system. J Neurol Neurosurg Psychiatry 1954; 17:216–224.
2. Kraus H. Therapeutic exercise. Springfield, IL: Charles C Thomas, 1949:58–60.
3. Fischer AA, Chang CH. Electromyographic evidence of paraspinal muscle spasm during sleep in patients with low back pain. Clin J Pain 1985; 1:141–154.
4. Kraus H. Use of surface anesthesia in the treatment of painful motion. JAMA 1941; 116(12):2582–2583.
5. Travell J, Simons D. Myofascial pain and dysfunction: the trigger point manual. Baltimore: Williams & Wilkins, 1983.
6. Kraus H. Clinical treatment of back and neck pain. New York: McGraw-Hill, 1970:119.
7. Kraus H, Eisenmenger-Weber S. Quantificative tabulation of postural evaluation. Physiother Rev 1946; 26: 235–242.
8. Lange M. Die Muskelharten (Myogelosen), Munich: JF Lehman, 1931.
9. Kraus H. Diagnosis and treatment of muscle pain. Chicago: Quintessence, 1988.
10. Simons DG. Myofascial pain syndromes due to trigger points. In: Goodgold J, ed. Rehabilitation medicine. St. Louis: CV Mosby, 1988:686–723.
11. Fischer AA. Pressure threshold meter: its use for quantification of tender spots. Arch Phys Med Rehabil 1986; 67:836–838.
12. Simms RW, Goldenberg DL, Felson DT, Mason JH. Tenderness in 75 anatomic sites. Arthritis Rheum 1988; 31:182–187.
13. Lautenschlager J. Bruckle W, Schnorrenberger CC, Muller W. Die Messung von Druckschmerzen im Bereich von Sehnen und Muskeln bei Gesunden und Patienten mit generalisierter Tendomyopathie (Fibromialgie-Syndrom). Z Rheumatol 1988; 47:397–404.
14. Wolfe F, et al. The American College of Rheumatology 1990 criteria for the classification of fibromyalgia. Arthritis Rheum 1990; 33:160–172.
15. Simons D. Muscular pain syndromes. In: Fricton JR, Awad EA, eds. Advances in pain research and therapy. New York: Raven Press, 1990:1–41.
16. Fricton JR. Myofascial pain syndrome: characteristics and epidemiology. In: Fricton JR, Awad EA, eds. Advances in pain research and therapy. New York: Raven Press, 1990:107–127.
17. Bennett RM. Myofascial pain syndromes and the fibromyalgia syndrome: a comparative analysis. In: Fricton JR, Awad EA, eds. Advances in pain research and therapy. New York: Raven Press, 1990:43–65.
18. Fischer AA. Pressure threshold measurement for diagnosis of myofascial pain and evaluation of treatment results. Clin J Pain 1987; 2:207–214.
19. Fischer AA. Application of pressure algometry in manual medicine. J Manual Med 1990; 5:145–150.
20. Fischer AA. Pressure algometry over normal muscles. Standard values, validity and reproducibility of pressure threshold. Pain 1987; 30:115–126.
21. Jensen K. Quantification of tenderness by palpation and use of pressure algometers. In: Fricton JR, Awad EA, eds. Advances in pain research and therapy. New York: Raven Press, 1990:165–181.
22. Reeves JL, Jaeger B, Graff-Radford SB. Reliability of pressure algometer as a measure of trigger point sensitivity. Pain 1986; 24:313–321.
23. Jaeger B, Reeves JL. Quantification of changes in myofascial trigger point sensitivity with the pressure algometer following passive stretch. Pain 1986; 27:203–210.
24. Graff-Radford SB, Reeves JL, Baker RL, Chiu D. Effects of transcutaneous electrical nerve stimulation on myofascial pain and trigger point sensitivity. Pain 1989; 37: 1–5.
25. Fischer AA. Pain and spasm alleviation by physiotherapy. Arch Phys Med Rehabil 1988; 69:736.
26. Fischer AA, Chang CH. Temperature and pressure threshold measurements in trigger points. Thermology 1986; 1:212–215.
27. Fischer AA. Documentation of myofascial trigger points. Arch Phys Med Rehabil 1988; 69:286–291.
28. Fischer AA. Clinical use of tissue compliance meter for documentation of soft tissue pathology. Clin J Pain 1987; 3:23–30.
29. Fischer AA. Objective measurements of muscle spasm. In: Jayson MIV, Swezey RL, Knoplich J, Hubault A, eds. Back pain, painful syndromes and muscle spasm. Current concepts and recent advances. NJ: Parthenon, 1990:65–69.

# Update on Postoperative Pain Management

Joel M. Kreitzer, M.D., Scott S. Reuben, M.D., and Allan P. Reed, M.D.

As a surveillance system for the prevention and detection of physically damaging events, pain may sometimes perform a useful service. Postoperative pain, however, usually serves no beneficial purpose. It may be associated with adverse physiologic effects, such as respiratory compromise from reflex splinting of thoracic and abdominal muscles; increased myocardial work and oxygen consumption; peripheral vasoconstriction; and adverse neuroendocrine effects, such as reduction in 5-day cumulative nitrogen balance (1). The psychological effects of postoperative pain include anxiety, fear, helplessness, and sleep deprivation, which alter the response to pain.

Despite widespread understanding of these effects, postoperative pain is often undertreated. A study of 109 postsurgical patients demonstrated a 75% incidence of moderate or marked painful states (2). Physicians and nurses have traditionally shared control over postoperative analgesia; consequently, their perceptions of pain management have an impact on the success of attempts to relieve pain. The major reasons for inadequate treatment of postoperative pain are fear of narcotic addiction and respiratory depression; inadequate knowledge of narcotic pharmacology; and a prevalent belief that complete pain relief is not the objective in treating postoperative pain (3).

In the past, patients assumed that pain would be a necessary component of their postoperative course. However, new modalities now allow patients to experience a relatively pain-free surgical course. The purpose of this article is to review these treatment options. Establishment of a dedicated postoperative pain service, led by anesthesiologists, has offered and maintained such treatment in large hospital centers (4).

## Neural Blockade

Within the past few years, new interest has arisen in the use of regional analgesic techniques for treatment of postoperative pain. Although modern technology has provided the means to safely and easily apply many of these techniques (interpleural analgesia, continuous brachial plexus analgesia), simple blockade of individual nerves or local areas can be very useful.

This concept has become especially useful in ambulatory surgery. During the repair of an inguinal hernia, blockade of the ilioinguinal and iliohypogastric nerves with a long-acting local anesthetic can provide excellent analgesia, lasting long into the patient's recovery period (5). This analgesia can be obtained without any motor or sympathetic blockade, allowing the patient to return home safely.

The same concept can be applied to blockade of nerves supplying the hands or feet. An ankle block performed with a long-acting anesthetic may supply up to 12 hr of excellent analgesia postoperatively. Simple injection of local anesthetic into a wound can also greatly decrease a patient's need for additional pain medications in the immediate postoperative period.

Blockade of intercostal nerves provides excellent pain relief after thoracic or upper abdominal procedures. Although complications such as local anesthetic toxicity, hypotension, and pneumothorax can occur, the incidence of serious side effects is low. This block can be performed multiple times during the postoperative period, or a cath-

---

From the Department of Anesthesiology, Mount Sinai School of Medicine (CUNY), New York, NY. Address reprint requests to Joel M. Kreitzer, M.D., Director, Pain Management Service, Department of Anesthesiology, Box 1192, The Mount Sinai School of Medicine, One Gustave L. Levy Place, New York, NY 10029-6574.

eter technique may be employed (6) for continuous intercostal blockade. In the future, interpleural analgesia may replace intercostal blocks.

Continuous blockade of the extremities is invaluable after vascular and orthopedic surgery to facilitate use of passive motion devices. Continuous axillary catheters used to anesthetize the brachial plexus are easy to insert and provide excellent analgesia with minimal side effects (7). The catheters may be injected many times, or a continuous infusion of a dilute local anesthetic may be administered for uninterrupted analgesia. After lower-extremity surgery, analgesia is provided with an inguinal catheter, which can be used for femoral nerve analgesia (8). Analgesia of the entire lower extremity can be obtained via a continuous inguinal perivascular block combined with a continuous sciatic block (9). These techniques provide analgesia without the motor or sympathetic interruption associated with continuous epidural analgesia.

Neural blockade may be accomplished for a longer duration by cryoanalgesia, in which nerves are temporarily frozen by application of a gas-expansion probe. Rapid cooling promotes formation of an ice ball, which destroys nerve tissue but leaves the nerve sheath intact so that regeneration may occur. Thus temporary analgesia, lasting weeks to months, may occur, followed by normal nerve regeneration. Although its use in postoperative pain care has not been firmly established, cryoanalgesia has been employed successfully after thoracotomy (10) and herniorrhaphy (11).

## Patient-Controlled Analgesia

Traditionally, narcotic analgesics are administered in the postoperative period by the intramuscular route. Inherent in such administration are many problems that reduce its efficacy: onset is slow, blood concentrations are variable, and delays in nurse response time may be 30 min or more (12). Large-bolus doses of intramuscular narcotics separated by several hours cause fluctuating blood levels, which lead to alternating periods of sedation and inadequate analgesia. The large-bolus method provides only short periods in which plasma levels attain the "therapeutic window." Plasma levels within the therapeutic window provide adequate analgesia without unwanted side effects (Fig. 1). Continuous intravenous infusions of meperidine achieve stable therapeutic blood levels which result in excellent pain control and minimal side effects (13). Thus, optimal postoperative analgesia can be achieved with methods that produce stable blood concentrations of analgesic drug. Although blood opioid concentration requirements may vary among patients, effective concentration requirements for the establishment of postoperative analgesia remain relatively constant for each individual.

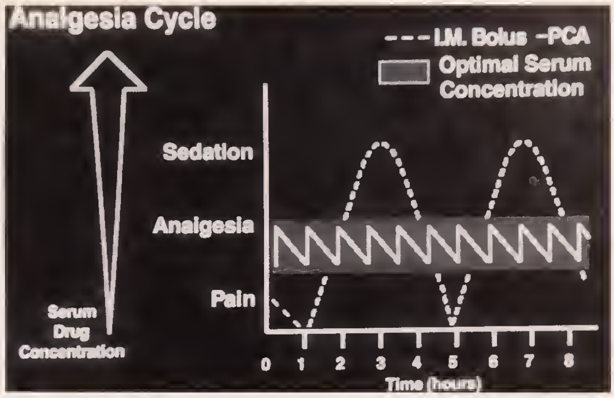


Fig. 1. Patient-controlled analgesia (PCA) delivers optimal serum concentrations of drug, providing adequate analgesia without unwanted side effects. *Reproduced* with permission from C. R. Bard, Inc., North Reading, MA.

Patient-controlled analgesia (PCA) is a technique that attempts to obtain these stable concentrations by frequent intravenous dosing of small amounts of narcotics. This approach originated about 30 years ago. In 1963, repeated intravenous injections of narcotics were recognized as providing more effective analgesia at lower total dosages than intramuscular administration (14). Unfortunately, this practice placed overwhelming demands on nursing personnel. This staffing problem was overcome in the early 1970s, when the first machines that enabled patients to self-administer narcotics (15, 16) were devised.

The basic unit (Fig. 2) consists of an infusion pump and a timing device. The pump accepts a syringe and tubing with a one-way valve connected to the patient's intravenous tubing. This mechanism allows for small, intermittent bolus injections on demand. To activate the system, the patient presses a button. Most devices also provide for continuous basal infusion, which helps to maintain constant blood levels of narcotic. The timer prevents additional intermittent injections within a predetermined time span called a "lockout" interval. Although the patient self-medicates, the physician retains control over the frequency of intermittent demand injections.

Most devices of this type display the total amount of drug administered and the number of times the button was pressed. Alarms to signal machine malfunction, intravenous tubing occlusion, or an empty syringe are available on many models. Unauthorized tampering with the narcotic syringe, dose, and time interval is prevented by a locking mechanism. The cost of a typical pump approximates $3,500.

To benefit maximally from the technique, patients should have a rudimentary understanding of the technique. The optimal time for instruction is preoperatively. Patients who expect complete pain relief will be dissatisfied with the results, because some discomfort is inevitable. Increased pain intensity is often caused by coughing, sneezing, walking, and wound redressing. When this pain is anticipated, an additional demand dose should be sought just prior to these events.

Application of the technique begins immediately after the operation, in the recovery room. Narcotics are administered intravenously until analgesia is achieved. Patients are then able to self-administer additional doses as blood levels fall and pain intensifies. If analgesia is subsequently unsatisfactory, the bolus dose is increased or the "lockout" interval decreased. If oversedation or respiratory depression occurs, the prescription is adjusted to provide less narcotic. Once a reasonable degree of analgesia is achieved, the patient may return to a regular nursing unit. Various opiates and dosage schedules are described elsewhere (17).

The patient and machine soon establish a fine-tuned feedback loop. When plasma levels fall and pain intensity increases, the patient demands more narcotic until better analgesia is provided. When plasma levels are excessive, causing sedation, the patient demands fewer additional doses until plasma levels fall again. This negative feedback loop helps prevent oversedation and respiratory depression. For this reason, the patients' visitors must not interfere. Misguided friends and family must not push the demand button for the patient, who they think is sleeping. Such "help" increases the risk of respiratory arrest.

The technique has many advantages in the treatment of postoperative pain. Bennett et al. (18–20) noted improved analgesia with decreased total narcotic consumption in patients undergoing gastric bypass surgery, compared to patients receiving intramuscular narcotics. Self-administered infusion avoids painful intramuscular injections, eliminates an unnecessary nursing function, and allows patients to participate in their own recovery process, a psychological advantage to many of them. Continuous analgesia promotes improvement in pulmonary function and may shorten hospital stays (21, 22). Recently, it has been suggested that there is less disturbance of nocturnal sleep during this therapy than with traditional pain therapy (23).

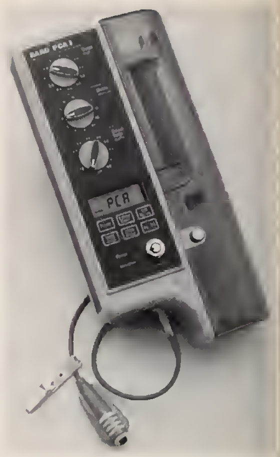


Fig. 2. Typical PCA pump with controls to regulate dosage, display to relate information, and push button by which patient requests medication. *Reproduced* with permission from C. R. Bard, Inc., North Reading, MA.

A major disadvantage of the technique is the cost of the machine, but cost may be justified in terms of more efficient use of nursing time and the possibility of reduced hospital stays. Clinically significant respiratory depression is rare. Even after major invasive surgery, normal arterial blood gas tensions are maintained during this therapy (24, 25). Cases of respiratory arrest have been described (26, 27), but they are few and thought to have been secondary to programming error or possibly machine malfunction. Tolerance to and withdrawal from narcotics are rare when this therapy is used for treatment of postoperative pain. Side effects of parenteral narcotics, such as nausea, vomiting, and sedation, are readily treated with routine standing orders.

Currently, most applications are directed toward intravenous administration of narcotics for postoperative pain relief. The techniques have recently been adapted to epidural methods of postoperative (28) and labor (29) analgesia. Even patients who lack intravenous and epidural access may benefit from this therapy. Subcutaneous self-administration for postoperative pain (30) and home treatment of chronic cancer pain (31) has been used in a limited way.

### Spinal Opiates

**Intrathecal and Epidural Routes.** The first application of spinal (intrathecal and epidural) opiates to humans was documented in 1979 (32). Since that time, spinal narcotics have been employed for treatment of postoperative pain, cancer pain, chronic pain, labor pain, and post-myocardial-infarction pain. Application of small amounts of narcotic close to the spinal cord receptors is thought to provide analgesia while limiting systemic side effects. Although this concept explains the action of intrathecal narcotics, confounding evidence exists regarding the action of epidural narcotics. Elevated plasma opiate concentrations are observed with the epidural use of lipid-soluble narcotics such as fentanyl (33). This implies a possible systemic contribution to the analgesia obtained from epidural opiates.

Mu, delta, kappa, and sigma are the four major classes of opiate receptors. Activation of mu receptors, located in the dorsal horn of the spinal cord, causes inhibitory modulation of afferent pain impulses entering the spinal cord from peripheral pain receptors. This "selective" analgesia occurs in the absence of sympathetic and motor blockade.

Whereas intrathecal opiates are administered directly into the cerebral spinal fluid (CSF) that bathes the spinal cord, epidurally administered opiates must first reach the CSF to exert their action. This is achieved primarily by diffusion of the un-ionized form of the opiate through the dura into the CSF. Consequently, the time course depends on the agent's lipid solubility. In the epidural space, fentanyl, a highly lipid-soluble opiate, has a short onset of action (4–10 min); the onset of action for morphine, a highly hydrophilic opiate, is much longer (20–30 min). The duration of action of epidurally administered fentanyl is short (approximately 4 hr) secondary to rapid diffusion out of the cerebrospinal fluid. Because of its hydrophilic nature, morphine remains in the CSF for a longer period of time and thus has a prolonged duration of action (18–24 hr).

Spinal opiates may be given either epidurally or intrathecally. The advantages of epidural opiates have been summarized (34) and include reduced incidence of side effects and prolonged use via an indwelling catheter. Advantages of intrathecal administration (35) include simplicity, reliability, and lower dose requirements. Selection of epidural versus intrathecal routes, as well as choice and dosing of opiates, is beyond the scope of this article, but is addressed elsewhere (36).

**Side Effects.** The limiting factors in the use of spinal opiates are side effects, which tend to occur more often after intrathecal than after epidural administration. Pruritus, especially on the head and face, is the most common side effect, occurring with an incidence of approximately 80% following intrathecal morphine (37) and 41% following epidural morphine (38) administration. The cause of pruritus remains elusive, but the condition is not related to increased histamine release (39). Its incidence can be decreased by employing more lipid-soluble opiates, such as fentanyl and continuous infusions, instead of intermittent dosing (40).

Nausea, vomiting, and urinary retention are the "minor" side effects related to spinal opiate use. Following morphine administered intrathecally or epidurally, they occur in approximately 50% and 40% of cases, respectively. These untoward effects occur less frequently with the more lipid-soluble narcotics. Low-dose intravenous naloxone can reverse the opiate-induced side effects without altering analgesia. Alternatively, untoward effects may be treated symptomatically.

Respiratory depression, a life-threatening situation, is the single most feared complication associated with spinal narcotic use. It occurs by two separate mechanisms. Initially, vascular uptake of the drug can cause high serum levels (33); late respiratory depression occurs secondary to cephalad spread of a hydrophilic opiate in the cerebrospinal fluid. Factors that predispose to development of respiratory depression include advanced age; use of water-soluble opiates; concomitant administration of intravenous opiates; increased intrathoracic pressure; large dosages; and use of the thoracic approach (41).

Although respiratory depression is rare, its true incidence remains unknown. The potential catastrophic results of this complication compel intense respiratory monitoring. Stay in an intensive care unit, private duty nursing, or, at the minimum 1- to 2-hr respiratory rate checks facilitate prompt diagnosis of respiratory impairment. Several approaches have been used to decrease the incidence of side effects. They include concomitant administration of low-dose intravenous naloxone (to abolish side effects yet retain analgesia); combinations of spinal morphine plus clonidine (42); and addition of opiate to dilute local anesthetic solutions (43).

**Benefits.** Spinal narcotics offer numerous important benefits with little risk to patients. Spinal narcotics provide improved pulmonary function (44–46) and superior analgesia compared to intravenous or intramuscular narcotics (47–49). In one study (50), patients preferred spinal narcotics over other parenteral routes for the treatment of postoperative pain. Intraoperative and postoperative use of epidural local anesthetics as well as of narcotics were compared to use of other parenteral narcotics (51). Patients receiving epidural narcotics demonstrated decreased mortality, less morbidity, shortened stays in intensive care unit, and earlier hospital discharges than those treated with intravenous narcotics.

### Interpleural Analgesia

Interpleural analgesia was first described by Kvalheim and Reiestad in 1984 (52). This new regional analgesia technique has been employed for pain management following cholecystectomy, thoracotomy, unilateral breast surgery, and renal procedures (53–58).

The safety and efficacy of interpleural analgesia, however, still remain controversial. Several technical problems associated with the blind insertion of interpleural catheters, as well as bupivacaine toxicity, have recently been reported (55). In addition, the analgesic efficacy of this technique for postthoracotomy pain management has recently been questioned (59).

As originally described by Kvalheim and Reiestad (52), an interpleural catheter is inserted at the completion of the surgical procedure, before the termination of general anesthesia. With the patient lying on the unoperated-on side, a 16-gauge Touhy needle is inserted at an angle of 30°–45° through the skin overlying the eighth intercostal space. The site of insertion is 10 cm from the midline spinous processes. The Touhy needle is advanced medially under the lower edge of the upper rib. After the intercostal membrane is perforated, the stylet is removed and a well-lubricated air-filled glass syringe is attached to the hub of the needle. The syringe and needle are advanced until a "click" is appreciated. The click represents entrance into the negative-pressure intrapleural space, which draws the syringe plunger forward. The syringe is removed, and a catheter is advanced 5–6 cm into the pleural space. During thoracotomy, the catheter is inserted under direct vision and is located in the dorsal pleural cavity at an intercostal space just above the incision.

The mechanism of action of interpleural analgesia still needs to be elucidated. According to one school of thought, local anesthetics diffuse from the pleural space into the intercostal nerve roots at the costovertebral angle. Others envision that local anesthetics diffuse out of the pleural space to the juxtaposed thoracic sympathetic chain, which carries afferent nociceptive fibers. The sympathetic chain, which lies within close approximation to the posterior thoracic cavity, can be effectively blocked as local anesthetics migrate out of the pleural space. A case report (60) documenting the development of Horner's syndrome following an interpleural injection illustrates the ability of this technique to block sympathetic neural activity.

The local anesthetic most frequently used for interpleural anesthesia is bupivacaine, with an average initial volume of 20 mL. Utilizing bupivacaine with epinephrine, Stromskag et al. (56) demonstrated that concentrations of 0.25%, 0.375%, and 0.5% provided equally effective pain relief. The duration of analgesia associated with each concentration was 285 min, 360 min, and 500 min, respectively. Brismar et al. (61), however, reported a significantly shorter duration of analgesia, lasting only 2–3 hr, following administration of either 0.25% or 0.5% bupivacaine after cholecystectomy.

To overcome the need for intermittent boluses of bupivacaine, several authors have attempted to infuse local anesthetic through an interpleural catheter (59, 60). Postthoracotomy pediatric patients received good analgesia from constant infusion of 0.5% bupivacaine (60), but postthoracotomy adult patients receiving constant infusion of 0.25% bupivacaine did not (59).

**Complications.** Since its introduction, interpleural analgesia has been associated with several complications. Syrmengt et al. (62) reported technical problems due to the blind insertion of interpleural catheters. These included intrapulmonary catheter placement with and without pneumothoraces. Two instances of tension pneumothoraces have been reported. Bupivacaine toxicity has also occurred in patients receiving interpleural anesthetics. Peak plasma concentrations of bupivacaine usually vary from 0.5 to 2.5 μg/mL, which approaches toxic levels (63). Utilizing the same dose of local anesthetic, Van Kleef et al. (64) reported significantly higher plasma bupivacaine concentrations following interpleural administration, compared to intercostal application. Seizure activity has been reported in one patient who demonstrated a plasma bupivacaine concentration of 4.9 μg/mL (55). In this case, a pleura inflamed by pneumonia was postulated to have absorbed local anesthetic unusually rapidly. Other potential side effects include phrenic nerve and thoracic sympathetic blockade.

Although originally heralded as a simple, safe, effective technique for control of postoperative pain, the interpleural administration of local anesthetics remains controversial and requires further evaluation.

### Acknowledgments

The authors thank Dianne Brown-Nelson for her editorial assistance.

### References

1. Cousins MJ. Management of postoperative pain. Int Anes Res Soc Review Course Lecture. 1986; 32–39.
2. Cohen FL. Postsurgical pain relief: patients' status and nurses' medication choices. Pain 1980; 9:265–274.
3. Weis OF, Sriwatanakul K, Alloza J, Weintraub M, Lasagna L. Attitudes of patients, house staff, and nurses

toward postoperative analgesic care. Anesth Analg 1983; 62:70–74.
4. Ready LB, Oden R, Chadwick HS, et al. Development of an anesthesiology-based postoperative pain management service. Anesthesiology 1988; 68:100–106.
5. Bugedo GJ, Carcamo CR, Mertens RA, Dagnino JA, Munoz HR. Preoperative percutaneous ilioinguinal and iliohypogastric nerve block with 0.5% bupivacaine for post-herniorraphy pain management in adults. Reg Anesth 1990; 15(3):130–132.
6. Safran D, Kuhlman G, Orhant EE, Castelain MH, Journois D. Continuous intercostal blockade with lidocaine after thoracic surgery. Clinical and pharmacokinetic study. Anesth Analg 1990; 70:345–349.
7. Gaumann DM, Lennon RL, Wedel DJ. Continuous axillary block for postoperative pain management. Regional Anesthesia 1988; 13:77–82.
8. Rosenblatt RM. Continuous femoral anesthesia for lower extremity surgery. Anesth Analg 1980; 59:631–632.
9. Smith BE, Fischer HB, Scott PV. Continuous sciatic nerve block. Anaesthesia 1984; 39:155–157.
10. Glynn CJ, Lloyd JW, Barnard JD. Cryoanalgesia in the management of pain after thoracotomy. Thorax 1980; 35:325–327.
11. Wood GJ, Lloyd JW, Evans PJD, Bullingham RES, Brittan BJ, Finch DRA. Cryoanalgesia and day-case herniorrhaphy (Letter). Lancet 1979; 2:479.
12. Vache I. Inadequate treatment of pain in hospitalized patients (Letter). N Engl J Med 1982; 55:307.
13. Stapleton JV, Austin KL, Mather LE. A pharmacokinetic approach to postoperative pain; continuous infusion of pethidine. Anaesth Int Care 1979; 7:25–32.
14. Roe BB. Are postoperative narcotics necessary? Arch Surg 1963; 87:912–915.
15. Forrest WH, Smethhurst PWR, Kienitz ME. Self-administration of intravenous analgesics. Anesthesiology 1970; 33:363–365.
16. Keeri-Szanto M, Heams S. Postoperative demand analgesia. Surg Gynecol Obstet 1972; 134:647–651.
17. White PF. Patient controlled analgesia: an update on its use in the treatment of postoperative pain. Anesthesiol Clin North Am 1989; 7(1):63–78.
18. Bennett RL. Patient controlled analgesia for the control of postoperative pain. Experience at the University of Kentucky. In: Harmer M, Rosen M, Vicker MD, eds. Patient-controlled analgesia. Oxford: Blackwell, 1985: 149–155.
19. Bennett RL, Batenhorst RL, Bivins BA, et al. Drug use pattern in patient-controlled analgesia (Abstract). Anesthesiology 1982; 57:A210.
20. Bennett RL, Batenhorst RL, Bivins BA, et al. Patient-controlled analgesia, a new concept of postoperative pain relief. Ann Surg 1982; 195:700–705.
21. Finley RJ, Keeri-Szanto M, Boyd D. New analgesia agents and techniques shorten postoperative hospital stay (Abstract). Pain 1984; 2:S397.
22. Walmsley PNH, Colclough GW, Mazloomdoost M, McDonnel FJ. Epidural PCA/infusion for post-nephrectomy pain: shorter hospitalization. (Abstract). Anesthesiology 1989; 71(3A):A684.
23. Owen H, Mather LE, Rowley K. The development and clinical use of patient controlled analgesia. Anaesth Int Care 1988; 16:443–447.
24. White WD, Pearce DJ, Norman J. Postoperative analgesia. A comparison of intravenous on demand fentanyl with epidural bupivacaine. Br Med J 1982; 2:166–167.
25. Tamsen A, Hartvig P, Fagerlund C, Dahlstrom B, Bondesson U. Patient-controlled analgesia therapy. Clinical experience. Acta Anaesth Scand 1982; 74:157–160.
26. White PF. Mishaps with patient-controlled analgesia. Anesthesiology 1987; 66:81–83.
27. Grey TC, Sweeney ES. Patient-controlled analgesia (Letter). JAMA 1988; 259:2240.
28. Boudreault D, Brasseur L. Epidural fentanyl by PCA and continuous administration (Abstract). Anesthesiology 1990; 73:A779.
29. Gambling DR, McMorland GH, Yu P, Laszlo C. Comparison of patient-controlled epidural anesthesia and conventional intermittent “top-up” injections during labor. Anesth Analg 1990; 70:256–261.
30. Taylor E, White PF, Urquhart ML. Postoperative analgesia: SQ-PCA vs IV-PCA (Abstract). Anesthesiology 1989; 71(3A):A685.
31. Campbell CF, Mason JB, Weiler JM. Continuous subcutaneous infusion of morphine for the pain of terminal malignancy. Ann Intern Med 1983; 98:51–52.
32. Wang JK, Wauss LE, Thomas JE. Pain relief by intrathecally applied morphine in man. Anesthesiology 1979; 50:149–151.
33. Loper KA, Ready CB, Downy M, et al. Epidural and intravenous fentanyl infusions are clinically equivalent after knee surgery. Anesth Analg 1990; 70:72–75.
34. Campbell C. Epidural opioids—the preferred route of administration. Anesth Analg 1989; 68:710–711.
35. Stoelting RK. Intrathecal morphine—an underused combination for postoperative pain management. Anesth Analg 1989; 68:707–709.
36. Gregg R. Spinal analgesia. Anesthesiol Clin North Am 1989; 7(1):79–100.
37. Abboud TK, Shnider SM, Dailey PA, et al. Intrathecal administration of hyperbaric morphine for the relief of pain in labor. Br J Anaesth 1984; 56:1351–1360.
38. Writer WDR, Hurtig JB, Edelist G, et al. Epidural morphine prophylaxis of postoperative pain: report of a double-blind multicentre study. Can Anaesth Soc J 1985; 32:330–338.
39. Korsh J, Ramanathan S, Parker F, Turndorf H. Systemic histamine release by epidural morphine (Abstract). Anesthesiology 1987; 67:A445.
40. Rutter DV, Skewes DG, Morgan M. Extradural opioids for postoperative analgesia: a double-blind comparison of pethideine, fentanyl, and morphine. Br J Anaesth 1981; 53:915–920.
41. Cousins MJ, Mather LE. Intrathecal and epidural administration of opioids. Anesthesiology 1984; 61:276–310.
42. Kitahata LM. Spinal analgesia with morphine and clonidine (Editorial). Anesth Analg 1989; 68:191–193.
43. Fischer RL, Lubenow TR, Liceaga A, McCarthy RJ, Ivankovich AD. Comparison of continuous epidural infusion of fentanyl-bupivacaine and morphine-bupivacaine in management of postoperative pain. Anesth Analg 1988; 67:559–563.
44. Bisback P, Rolly G, Tampabolon O. Efficacy of the extradural administration of lofentanil, buprenorphine, or saline in the management of postoperative pain. Br J Anesth 1985; 57:943–948.
45. Bromage PR, Camporesi E, Chestnut D. Epidural narcotics for postoperative analgesia. Anesth Analg 1980; 59:473–480.
46. Rawal N, Sjostrand U, Dahlstron B, Mydahl P, Ostelius J. Epidural morphine for postoperative pain relief: a comparison study with intramuscular narcotic and intercostal nerve block. Anesth Analg 1982; 61:93–98.
47. Reiz S, Ahlin J, Ahrenfeldt B, Anderson M, Anderson S.

Epidural morphine for postoperative pain relief. Acta Anaesth Scand 1981; 25:111–114.

48. Raj PP, Knarr D, Vigdorth E, et al. Comparison of continuous epidural infusion of a local anesthetic and administration of systemic narcotics in the management of pain after total knee replacement surgery. Anesth Analg 1987; 66:401–406.
49. Welchew EA, Thornton JA. Continuous thoracic epidural fentanyl. A comparison of epidural fentanyl with intramuscular papaveretum. Anaesthesia 1982; 37:309–316.
50. Brownridge P, Frewin D. A comparative study of techniques of postoperative analgesia following cesarean section and lower abdominal surgery. Anaesth Int Care 1985; 13:123–130.
51. Yeager MP, Glass DD, Neff RK, Brinck-Johnsen T. Epidural anesthesia and analgesia in high-risk surgical patients. Anesthesiology 1987; 66:729–736.
52. Kvalheim L, Reiestad F. Interpleural catheter in the management of postoperative pain (Abstract). Anesthesiology 1984; 61:A231.
53. Reiestad F, Stromskag KE. Interpleural catheter in the management of postoperative pain, a preliminary report. Reg Anesth 1986; 11:89–91.
54. Reiestad F, Stromskag KE, Holmquist E. Interpleural administration of bupivacaine in postoperative management of pain (Abstract). Anesthesiology 1986; 65:A204.
55. Seltzer JL, Larijani GE, Goldberg ME, Marr AT, Rocci ML. A kinetic and dynamic evaluation of intrapleural bupivacaine for subcostal incisional pain (Abstract). Anesthesiology 1986; 65:A213.
56. Stromskag KE, Reiestad F, Holmquist E. Intrapleural administration of 0.25%, 0.375%, and 0.5% bupivacaine with epinephrine after cholecystectomy. Anesth Analg 1988; 67:430–434.
57. Reiestad F, Stromskag KE, Heen L. Intrapleural catheter for analgesia after thoracotomy in adult patients (Abstract). Acta Anaesth Scand 1987; 31(Suppl):A237.
58. Kambam JR, Handte RE, Flanagan J, Fisher K, Lupinetti M, Hammon J. Intrapleural anesthesia for post-thoracotomy pain relief (Abstract). Anesth Analg 1987; 66:S90.
59. Rosenburg PH, Scheinin BMA, Lepantalo MJA, Lindfors O. Continuous intrapleural infusion of bupivacaine for analgesia after thoracotomy. Anesthesiology 1987; 67:811–813.
60. McIlvaine WB, Knox RF, Fennessey PV, Goldstein M. Continuous infusion of bupivacaine via intrapleural catheter for analgesia after thoracotomy in children. Anesthesiology 1988; 69:261–264.
61. Brismar B, Pettersson N, Tokics L, Strandberg A, Hedenstierna G. Postoperative analgesia with intrapleural administration of bupivacaine-adrenaline. Acta Anaesth Scand 1987; 31:515–520.
62. Syrmengt, Gomez MN, Johnson B, Rossi N, Chiang CK. Intrapleural bupivacaine—technical considerations and intraoperative use. J Cardiothorac Anesth 1989; 3:139–143.
63. Tucker GT. Pharmacokinetics of local anesthetics. Br J Anaesth 1986; 58:717–721.
64. Van Kleef JW, Burm AGL, Vletter AA. Single-dose interpleural versus intercostal blockade: nerve block characteristics and plasma concentration profiles after administration of 0.5% bupivacaine with epinephrine. Anesth Analg 1990; 70:484–488.

# Postoperative Pain Management in Children

Charles M. Haberkern, M.D., Donald C. Tyler, M.D., and Elliot J. Krane, M.D.

Over 3 million children undergo surgery each year in the United States (1, 2). The available data indicate that pain management in children is often inadequate and that children receive proportionately less postoperative analgesia than adults for comparable surgical procedures (3–5). One might wonder why this is so, when relief of pain is one of the goals of medicine.

Based on studies of adults, one may infer that a significant part of the problem lies in the routine administration of a fixed dose of opioid, ordered to be given as needed by intramuscular injection. Similar pain management regimes are used in children and contribute to the inadequacy of their postoperative pain management. Imagine what goes through a child's mind when the nurse approaches with the instrument most often feared by children—a syringe and needle—and tells the child that this shot will produce relief of pain. The nurse gives the shot, the child suffers the pain of injection, and, not surprisingly, does not feel better. Later, when the narcotic begins to take effect, the child is not convinced that the injection had anything to do with relief of pain. This scenario, when analyzed for its behavioral components, reveals a situation in which the child's complaint of pain is followed by an aversive stimulus; one would thus expect the child to reduce the frequency of pain complaints. That result makes things quieter on the ward for the nursing staff, but the child suffers pain quietly and, as a result, receives inadequate treatment. Such a phenomenon unfortunately occurs in many hospitals.

In addition to the routine use of a fixed prescription of intramuscular opioid, other factors contribute to inadequate pain treatment in children after surgery. One is the incorrect belief that infants and children do not feel pain or, if they do, do not remember it. There is no credible evidence that children do not feel or remember pain and, in the absence of evidence, we are certainly compelled to treat pain. Another contributing factor is lack of specific knowledge about the assessment, physiology, pathophysiology, and treatment of pain in children. Unfounded concerns about side effects of opioids and about narcotic addiction contribute to undertreatment of children after surgery.

In this article, the physiology and pharmacology of pain transmission, the response of the patient to pain, the effects of untreated pain, and pain assessment are addressed. We then discuss the unique problems of pain management in infants, and we conclude with a general discussion of techniques of pain control.

## Physiology and Pharmacology

Pain is a perception that is far more complex than simple transmission of information from a site of tissue damage along nerve pathways to the brain. Pain consists of a component of transmission of pain sensation, a component of processing and evaluation by higher centers of the brain, and a component of reaction to the sensation. Although the pathways of nerve transmission to the brain are complex in themselves, it is the evaluation, processing, and reaction to pain sensation that are the major source of complexity and that account for the wide variability in response to pain among patients. To try to identify some of

---

From the Multidisciplinary Pain Program, Departments of Anesthesiology and Pediatrics, Children's Hospital and Medical Center, University of Washington School of Medicine, Seattle, Washington. Address reprint requests to Charles M. Haberkern, M.D., Department of Anesthesiology, Children's Hospital and Medical Center, P.O. Box C5371, Seattle, WA 98105.

the specific problems associated with pain in children, we briefly discuss each of these components, namely, pain transmission, the psychological components of processing and evaluation, and the responses to pain.

**Pain Transmission.** In the periphery of the nervous system are located nociceptors that respond to pain and inflammation. The nociceptors are specialized units that respond to strong pressure, heat, cold, irritant chemicals, or other stimuli. Some mediators of inflammation, such as prostaglandins, bradykinin, 5-hydroxytryptamine, histamine, substance P, and acetylcholine, can serve as excitors or modulators of the painful stimulus. There is not a great deal of information on the changes that occur in these systems during human growth and development, although there is evidence that cutaneous sensory perception exists in the fetus and that the density of cutaneous nociceptive receptors in newborns is equal to or greater than that in adults (6). The painful sensations that arise in the periphery are transmitted to the spinal cord via A delta and C nerve fibers (small myelinated and nonmyelinated fibers), which enter the substantia gelatinosa of the spinal cord. Here many synapses occur, along with modulation of the pain sensation, both from descending nervous input and from other peripheral input. Modulation may also result from the effects of chemical mediators, such as the endogenous opioid system, or from the psychological influences that modulate pain. It is in the area of the substantia gelatinosa that exogenous opioids are felt to have a major effect in reducing the perception of pain. After the pain sensation passes through the substantia gelatinosa, it ascends to the brain via the anterior spinothalamic tract and other ascending tracts. In the brain, the sensation traverses the thalamus and projects to the cerebral cortex.

**Psychological Components.** The psychologic influences that alter a patient's response to pain are complex. For an adult, the context of the pain may alter the perception of pain. For a child, chronological age, developmental stage, cognitive abilities, behavioral patterns, expectations, and family interactions are major factors that influence the response to pain; the child's trust of caregivers and previous painful experiences may also influence response. In addition, all these factors may affect learning about pain and the way the child deals with pain in the future.

**Responses to Pain.** The response to pain consists of behavioral, psychological, and social changes, which are discussed below, and of physiologic changes that may alter the pain response itself. Segmental reflexes arise at the level of the spinal cord and may lead to muscle spasm, itself causing more pain, and to alterations of respiration, circulation, and other organ function. Some of these changes may arise from activation of the sympathetic nervous system. Suprasegmental reflexes arise from stimulation of higher medullary or autonomic centers, and alterations of respiratory or circulatory function may thereby be produced. In addition to these reflex responses, pain can produce an endocrine stress response, leading to an increase in secretion of epinephrine, norepinephrine, ACTH, cortisol, growth hormone, antidiuretic hormone, glucagon, and aldosterone, and to a decrease in secretion of insulin and testosterone (7). If unmodulated, the stress response may lead to complications that slow a patient's recovery.

## Effects of Untreated Pain

Untreated pain can lead to physiologic complications, psychologic distress, family disruption, interruption of hospital routine, and prolongation of hospitalization with resultant increased costs. It may be hard to separate all the harmful effects of hospitalization and surgery from those of pain itself in any individual child, but inadequately treated pain can clearly color a child's hospital experience and may account for many adverse effects. We discuss each of these effects of untreated pain.

**Physiologic Complications.** It is clear that adverse physiologic consequences arise in association with untreated pain. Following upper abdominal surgery or thoracic surgery, pain can lead to splinting, ineffective coughing, and potentially to hypoxia, atelectasis, and pneumonia (8–10). Data are lacking in children, but studies in adults indicate that, after upper abdominal surgery, there is a decrease in functional residual capacity (FRC) (11, 12), a decrease in inspiratory and vital capacity, and a fall in arterial oxygen tension (13). Three factors in children may accentuate these problems. First, young children have more compliant chest walls (14, 15), require muscle activity to stabilize the chest wall during inspiration, and, therefore, may be more prone to respiratory muscle fatigue. Second, closing capacity is greater than FRC in young children and infants (16), so a fall in FRC makes a child particularly susceptible to hypoxemia. Finally, the higher oxygen consumption rate and the higher alveolar ventilation:FRC ratio in a young child may predispose the child to hypoxemia associated with pain in the postoperative period. Adequate pain therapy will help the child to cough, take deep breaths, and reduce the harmful effects of surgery, anesthesia, and pain on respiratory function. Activation of the endocrine stress response leads to an increase in catecholamine secretion and, therefore, to an increase in heart rate, blood pressure, systemic and pulmonary vascular resistance, and oxygen consumption. If catecholamine secretion continues at a high rate, it can lead to a reduction in caloric reserve needed for healing.

**Psychologic Distress.** Psychologic distress may arise in a child just from being in a hospital and separated from parents, and pain can certainly add to this distress. The nature of the response will depend on the age of the child and the other factors affecting the response to pain mentioned above. Pain leads to anxiety and a loss of a sense of control, which can be particularly harmful in a patient who is at a developmental stage in which he is learning self-control, i.e., in the early school-age child. The child's social relationships may also be harmed: in young burned patients, a condition of "learned helplessness" has been described in which children withdraw from their environment and stop participating in interpersonal interactions (17). In addition to social withdrawal, behavioral regression, temper tantrums, demanding behavior, and other reactions to pain may occur, which may have long-term effects on the way a child learns to deal with the world.

**Disruption of Family and of Hospital Routine.** Family disruption may occur. Families are under considerable stress with a child in the hospital, and if the child also has inadequately treated pain, the stress on the family may increase and magnify any tension and social pathology already present.

For the hospital staff, a child in pain presents a problem. On the one hand, the bedside nursing staff's duty to patients requires that the child be treated, and the staff is frustrated if those in charge of medical care are unwilling to, or incapable of, adequately treating the pain. On the other hand, physicians may be frustrated because of a desire to treat pain accompanied by a perception of a lack of safe techniques to do so. Inordinate amounts of time may be expended discussing the untreated pain, and other factors in the patient's care may thereby be neglected. In addition, the crying or moaning of a patient in pain can upset the hospital routine and produce distress in other patients and families.

**Prolongation of Hospitalization.** All these factors can lead to prolonged hospitalization and increased costs. A recent study compared a group of adult patients who had anesthesia with one technique (combined epidural with "light" general anesthesia, and postoperative pain control with epidural local anesthetics or opioids) to a group of patients treated with conventional general anesthesia and postoperative pain control with parenteral opioids (18). The group that received epidural anesthesia and epidural postoperative pain control had fewer postoperative complications and reduced hospital costs. Although the results cannot be attributed entirely to better postoperative pain control, it is suggested that better pain control had a major influence on outcome.

## Pain Assessment

There are no physiologic or laboratory methods of measuring pain intensity in children or adults. And, because of the developmental, cognitive, and emotional differences between adults and children, pain assessment is even more difficult in children. Generally, two types of techniques are useful for pain assessment: self-report and observation of behavior. Variations of these techniques are useful for estimating pain intensity in children.

**Self-Report.** The simplest, most accurate method of self-report of pain intensity in adults and older children is the Visual Analog Scale (VAS). The scale is an undivided 10-cm line which has the words "no pain" placed on one end and the words "worst pain possible" on the other. The patient is instructed to make a mark on the line, the position of the mark indicating the intensity of the pain. A variation is a scaling technique in which the patient estimates pain intensity by selecting a number from zero to ten, zero indicating no pain. Use of this technique presupposes patient alertness and an understanding of the number system, or at least an understanding that one quantity can consistently be greater than another. The VAS and its variations have proven useful with children over the ages of six or seven years, and reasonable validation data are available for this group (19–21).

Pain assessment by self-report is more difficult in patients under age six years because they usually have not fully developed the cognitive skills necessary for use of the VAS or a numerical scaling system, and they usually have not fully developed the language skills to effectively communicate needs and desires. In addition, the methods of self-report in this age group rely on cooperation of the child, which may be lacking in the strange setting of the hospital after a painful operation and which may be affected by other factors, including hunger, cold, fatigue, anxiety, and separation from parents. Among the methods of self-report are the Smiley Analog Scale, the Poker Chip scale, the Pain Ladder, the Oucher Scale, and the color scale. The Smiley Analog Scale, in which a child is asked to select from a series of pictures of faces the one face that depicts how he or she feels at the moment, has provided investigators with some useful information about pain control in young children (20, 22).

**Observation of Behavior.** Behavior observation is most often used in the preverbal child and the verbal child who is unable to self-report. One method of behavioral observation is the CHEOPS Scale, for which an observer notes six aspects of the child's behavior and applies a numerical rating to each, the total score estimating pain intensity. This scale has been tested in postoperative children and has been found a useful measure of presumed pain (23).

We have used a simple 5-point estimation of behavior as a guide for determining need for postoperative pain medication and have found it useful (24), but validation studies are not yet completed. We score 1 point for a child who is laughing and playing; 2 points for a child who is happy; 3 for one who is neutral; 4 for crying or other behaviors indicating pain from which the child can be distracted; and 5 if the child cannot be distracted. Pain medication is administered if the child scores 4 or 5.

Other structured observational scales are also available (25), and all serve to improve on the often inadequate management of pain based on unstructured observation of a child by the staff.

For the reader who is interested in further discussion of pain assessment, several recent reviews are available (26–29).

## Treatment of Postoperative Pain in Infants

Infants and adults have historically been treated differently because it has been assumed that infants do not feel pain or, if they do, that they do not remember it. These attitudes are being challenged as evidence evolves that even newborn infants have the anatomic and functional components needed for pain perception, that they do "feel pain," and that harmful physiologic effects arise as a result of untreated pain (30). Clearly, the areas of the higher nervous system necessary for conscious perception of pain are present in newborns, and positron emission tomography scans have shown that these higher centers are active (31). The fact that facial expressions characteristic of the emotional responses to pain are produced in young infants is further evidence that higher nervous system centers are functioning. That habituation and operant conditioning occur is evidence that memory is also present in infants (32). Studies of infants undergoing surgery with very "light" anesthesia indicate that surgical stimulation produces endocrine stress responses, and that these stress responses lead to an extended catabolic period following surgery and a higher number of complications than are seen with deeper anesthesia (33).

Given that infants do feel pain, that pain affects their behavior, and that the stress response indicates that there are valid physiologic reasons to treat pain, one might ask what obstacles stand in the way of effective treatment of postoperative pain in infants. The two primary obstacles are the inadequate methods available to assess pain and the concern for increased risks associated with parenteral opioids in the infant.

Pain assessment is a major challenge because we must rely on subjective evaluation of the infant's behavior. Although several behavioral scoring systems are now available, including a 20-point behavior score by Attia et al. (34), they are not well standardized or validated in infants. Attempts have been made to quantitatively describe the type of cry that a baby emits (35), but it is difficult to correlate any particular type of cry with pain unless the context of the situation is taken into account. Investigators have attempted to evaluate facial expression (36), and although it is clear that facial expression does correlate with presumed pain, determination of intensity is difficult. Thus, physicians and nurses who are assessing pain in infants usually rely on intuition and clinical experience rather than on an objective pain-scoring system.

The other major impediment to treating infant pain after surgery is the concern for safety. Opioids are potentially hazardous to young infants and, in fact, major textbooks suggest avoiding opioids unless the babies are to be mechanically ventilated and closely monitored. Newborn infants have been shown to have decreased ventilatory responsiveness to carbon dioxide (37) and greater respiratory depression induced by opioids (38) than do older children and adults. In addition, morphine clearance has been shown to be prolonged in the first week of life (39), and if data from infant rats can be extrapolated (40), the infant's blood-brain barrier is more permeable to opioids than the adult's. At the same time, kinetic studies on morphine and the potent synthetic narcotics indicate that clearance and elimination half-life in the 2- to 3-month-old infant approximate those of the older child and adult (41). Moreover, it has recently been shown that ventilatory depression induced by intravenous fentanyl, a potent opioid, is no greater in the infant older than three months than it is in the child and adult (42). Therefore, available data suggest that opioid administration for treatment of postoperative pain may be much safer in the infant older than three months than was previously understood.

In light of the difficulties in pain assessment and concerns for the safety of opioid administration in infants, only general suggestions can be given for postoperative pain management in these patients. If an infant is to be mechanically ventilated after surgery, the solution to the problem of treating postoperative pain becomes obvious: use opioids in sufficient doses to abolish physiologic and behavioral manifestations of pain (morphine sulfate bolus or infusion therapy, as described below) and, when appropriate, wean the infants from the ventilator and opioid therapy with caution. If the baby is not mechanically ventilated, one may administer acetaminophen in doses of 10–15 mg/kg every 4–6 hr (orally or rectally). If this does not provide sufficient pain relief, one may add codeine (1 mg/kg every 4–6 hr) or cautiously use morphine in low doses (0.05 mg/kg every 3–4 hr) with careful monitoring for respiratory depression. Regional anesthetic techniques, described briefly below, may also be used in the infant, especially for treatment of pain following circumcision and herniorrhaphy.

### Treatment of Pain in Children over 6 Months of Age

**Opioids.** The mainstay of treatment of postoperative pain in children is parenteral opioids. The undesirability of the intramuscular route in children has prompted attention to alternatives to that traditional approach. Sublingual, nasal, rectal, and transdermal routes are being investigated, but appropriate preparations are generally not yet available for children. Therefore, the intravenous route is preferred if adequate nursing observation and monitoring are available and if sufficient doses are given at appropriate dose intervals to provide effective pain relief.

The choice of drug generally depends on clinical experience and, ideally, objective information concerning the various drugs. Morphine has the advantage of great clinical experience and economy, and it is the standard against which all other opioids are judged. Starting doses for children are 0.05–0.10 mg/kg intravenously, given every 2–4 hr, although considerably higher doses may be needed and safely tolerated by certain patients. The side effects associated with morphine are those commonly associated with opioids: respiratory depression, sedation, urinary retention, nausea, and pruritus. Meperidine, a synthetic opioid, is a frequently used alternative for postoperative analgesia, but there is insufficient scientific data to compare the incidence of side effects between the two drugs in children. Morphine is metabolized (glucuronidated) by the liver, and the metabolic by-products are excreted by the kidneys. Meperidine is metabolized to nor-meperidine, a centrally active compound excreted by the kidneys. Accumulation of meperidine metabolites may lead to central nervous system excitation manifested as delirium, agitation, or seizures. For this reason, meperidine should be avoided in patients who might accumulate the metabolites, specifically patients with renal impairment.

Other synthetic opioids are also available, including fentanyl, sufentanil, and alfentanil. These drugs are more receptor specific and, therefore, have the potential advantage of providing better analgesia with fewer side effects. At the same time, they are short-acting, expensive, and, therefore, not well suited to providing postoperative pain relief via intermittent, parenteral administration. Fentanyl, the most commonly used of these agents, can be given as a continuous infusion for postoperative pain relief in a closely monitored setting (see below). All three drugs have minimal cardiovascular side effects and do not cause histamine release; however, they may compromise ventilation by depressing respirations and producing chest wall rigidity.

Because some physicians continue to use only intramuscular injections for pain relief, investigators have searched for medications that can minimize the number of injections used. Methadone has long-lasting analgesic effects (43); administration (0.2 mg/kg) at the start of anesthesia provides postoperative analgesia for up to 36 hr. The use of this medication therefore obviates the need for frequent intramuscular injections for a period after surgery, and it may be continued later with oral administration (0.3–0.4 mg/kg per day). The potential hazard of methadone is, of course, that an actual or relative overdose is also long-lasting.

Codeine is sometimes used to provide moderate analgesia when the physician wishes to avoid other potent opioids or to use an agent that is effective orally (0.5–1.5 mg/kg every 4–6 hr). The analgesic efficacy of codeine is additive to the effects of aspirin and acetaminophen. Side effects with codeine have not been well studied in children, but may include nausea and vomiting and, with intravenous use, flushing and hypotension, possibly related to histamine release (44).

**Nonopioids.** Nonopioids include acetaminophen, aspirin, and the nonsteroidal anti-inflammatory drugs (NSAIDs).

Acetaminophen is effective for relief of mild to moderate pain in both adults and children. Children's preparations are available for both oral and rectal administration, and they are frequently used for pain relief in children of all ages (dose 10–15 mg/kg every 3–4 hr). There are no major side effects; however, massive overdose, if untreated, can lead to fatal hepatic necrosis.

Aspirin is effective in providing pain relief, but it has side effects and possible toxicity that now limit its use in children following surgery. Aspirin inhibits platelet aggregation and may lead to impaired coagulation. In addition, the use of aspirin has been associated with the development of Reye's syndrome (45), a potentially fatal liver dysfunction of children and adolescents.

NSAIDs relieve mild to moderate postoperative pain in adults, and they may be used in children (for example, ibuprofen 5–10 mg/kg orally every 4–6 hr, naproxen 5 mg/kg every 12 hr). Except for the recently released drug ketorolac, no parenteral preparations or satisfactory rectal preparations are available, so the use of NSAIDs is generally limited to those children who are able to take medications orally. We are aware of no studies that have evaluated the safety and efficacy of ketorolac in children.

**Behavioral Treatment.** Behavioral treatment involves providing both a comfortable environment for all children during the surgical period and active behavioral measures that may be useful for older children.

Good preoperative preparation of the child is important. He or she should be informed in a nonthreatening way about what will happen during the surgical period, that separation from parents will be minimized, and that pharmacologic pain relief is available if needed. One educational approach is the use of puppets or dolls at a formal preoperative preview of the hospital stay. If possible, specific fears (such as the child's equating being put to sleep with what may once have happened to a family pet) need to be identified and discussed. Minimizing separation of the child from the family is helpful in reducing many of the anxieties that can increase a child's perception of pain.

With respect to behavioral pain management techniques, distraction techniques, such as listening to headphones or stories told by staff or parents, can be helpful. One active technique that can be taught to a child is progressive muscular relaxation. The rationale for this approach is that pain and anxiety may lead to muscle tension that can amplify pain. The child is taught to produce deep muscle relaxation prior to surgery and then to use this technique to reduce pain after surgery. Meditation, mental imagery, and self-hypnosis are frequently useful in producing relaxation in children, who can be excellent hypnotic subjects. The child can be taught to develop a mental image of comfort in the affected part of the body: for a painful limb, the suggestion might be made to think of the limb immersed in a cold stream or floating free on a cloud.

## New Techniques of Pain Control

Recent advances in pain management have improved pain control for adults and can to some extent be applied in children. Examples are continuous intravenous infusion of opioids, patient-controlled analgesia (PCA), epidural administration of opioids, and regional anesthesia. These techniques are discussed briefly, emphasizing their use in children.

**Continuous Intravenous Infusion.** To avoid the fluctuating blood levels that occur with intermittent use of parenteral opioids, continuous intravenous infusion of opioids can be used. Prerequisites for this technique are close nursing coverage to watch for signs of excess sedation, and a reliable method of providing a constant opioid infusion rate. Our clinical approach is to use a loading dose of opioid sufficient to achieve a therapeutic serum level to make the child comfortable (for example, morphine 0.05–0.10 mg/kg, or more if necessary), and then to begin a continuous infusion of opioid (for example, morphine, 0.01–0.02 mg/kg per hr) adjusted according to patient need. (The infusion may be easily prepared by mixing the number of milligrams of morphine equal to the patient's weight in kilograms in 100 mL of diluent; the infusion is then administered at 1–2 mL/hr; e.g., for a 20-kg child, a 100-mL solution with 20 mg morphine infused at 1 mL/hr provides 0.01 mg/kg per hr of morphine.) Infusion rates less than 0.02 mg/kg per hr are not associated with carbon dioxide retention (46). Our further observation is that tolerance to an infusion rate of morphine often occurs after 48–96 hr of therapy, requiring an increase in the infusion rate.

Often we find the need to supplement a continuous intravenous infusion of opioids with benzodiazepine therapy (either bolus or continuous) to reduce muscle spasms following certain procedures or to reduce anxiety. The relatively short-acting agent midazolam (0.01–0.10 mg/kg/hr) is particularly useful for continuous therapy. Although benzodiazepines work synergistically with opioids to relieve postoperative pain and discomfort, they also potentiate opioid side effects, especially sedation and respiratory depression, and therefore call for close monitoring of the patient. Adjunct therapy with benzodiazepines may also be useful with the other techniques of pain control described below.

**Patient-Controlled Analgesia.** Patient-controlled analgesia (PCA) is a technique of opioid administration based on the premise that a patient is the best judge of how much analgesic he or she needs, and how often it should be given. The technique uses a commercially available, sophisticated intravenous pump that allows the patient to self-administer an opioid within safe limits determined by the physician. When the patient pushes a button, the intravenous pump delivers a small bolus of analgesic; the pump then goes into a lockout period during which the patient cannot administer further doses and during which the full effect of the last dose can be experienced. The physician determines the size of the bolus, the duration of the lockout interval, and an hourly or four-hourly maximum dose. For a typical adult patient, the bolus dose is 1–2 mg of morphine, the lockout interval is 6–10 min, and the 4-hr limit is 15–25 mg. This method allows the patient to control his or her own analgesic use and to treat pain when it occurs, not according to a preset hospital schedule. The use of small frequent boluses of opioids allows the patient to maintain a more constant therapeutic blood level, thus avoiding the cycle of sedation-analgesia-pain so commonly seen with bolus parenteral opioid administration. In addition, most of these devices have the capability of providing a continuous background infusion of opioid (e.g., morphine, 0.01–0.03 mg/kg per hr), which is particularly useful in the control of pain during the first postoperative day.

We have reported our experience with PCA n 25 adolescent patients following surgery (47). The patients, aged 12–19 years, used the machine effectively, no overdoses occurred, and the subects were satisfied with their analgesia. Those patients who had had previous surgery uniformly preferred this type of analgesia to conventional echniques. Medication patterns of opioid use vared widely among patients (as much as eightfold, orrected for body weight), and most patients did not titrate themselves to complete analgesia but, rather, to a slight degree of pain. There were no episodes of clinical respiratory depression.

We now routinely offer PCA to patients 8 years old and older if they understand the concept of PCA, wish to use the device, and have no prior history of drug abuse. In the postoperative recovery room, the patients are medicated to comfort by an anesthesiologist with intravenous opioids, and PCA is begun while the patients are still comfortable. Typical orders with morphine are: dose, 0.01–0.02 mg/kg; lockout, 6–10 min; and 4-hr limit, 0.20–0.40 mg/kg. In each instance, a member of the anesthesiology department is identified to respond to problems, to adjust doses, and to visit the patient daily to evaluate pain control and side effects.

**Epidural Opioids.** Administration of opioids into the epidural space has achieved widespread use in adults and, more recently, in children. (Intrathecal opioids have been used in children only to a limited extent [48].) The technique makes use of the fact that opioid receptors are found in the substantia gelatinosa of the spinal cord, and administration of opioids into the epidural space leads to occupation of these receptors by the drug and a subsequent decrease in pain transmission. The pain relief obtained is not sufficient to provide surgical anesthesia, but it is sufficient to make most patients quite comfortable during the postoperative period. In addition to good-quality pain relief, advantages of epidural opioids include freedom from impaired motor function and normal function of the sympathetic nervous system. Moreover, since analgesia can be achieved with lower doses of opioid by the epidural route than by most other routes, serum drug concentrations are lower and, therefore, the risk of serious side effects, such as respiratory depression, is lower.

Epidural narcotics have been used extensively for children. Pain relief has been shown to be longer lasting than with either intravenous morphine or epidural bupivacaine (24). Our indication for epidural narcotics is expected postoperative pain of moderate to severe intensity, e.g., pain following thoracotomy, laparotomy, genitourinary surgery, or lower-extremity orthopedic surgery. A catheter is usually placed in either the lumbar epidural space or the caudal space to allow repeat doses. In children younger than 8 years of age, the caudal space is particularly simple to locate and provides reliable results. An epidural catheter may be placed prior to surgery, to provide regional anesthesia intraoperatively, or, much more commonly in children, after the induction of anesthesia, to provide a local anesthetic agent to supplement general anesthesia.

We prohibit concomitant parenteral narcotics during administration of the anesthetic to reduce the risk of postoperative respiratory depression, and we administer epidural morphine at least one-half hour prior to the end of surgery to allow time for the onset of analgesia to occur. We use a dose of 0.030–0.075 mg/kg of preservative-free morphine, the exact dose depending on the site and nature of the surgical procedure. Subsequent doses can be adjusted based on the degree and duration of analgesia and on the presence of side effects (49).

Preservative-free morphine is the drug most commonly used for epidural analgesia (Duramorph, Elkins-Sinn, Cherry Hill, NJ; Astramorph, Astra Pharmaceutical Products, Westboro, MA). Some investigators have suggested that a more lipophilic drug, such as fentanyl or sufentanil, might be preferable to morphine, because less of such an agent might remain in the cerebrospinal fluid and ascend to the brain to produce respiratory depression and other side effects. Fentanyl has been used in adults, but data are not available to support the hypothesis that side effects are reduced in children. Fentanyl does have the advantage that it has a shorter onset of action than morphine, but fentanyl also has a shorter duration of action, which makes bolus dosing less effective. We are currently investigating the use of epidural sufentanil for postoperative analgesia in children.

Side effects of epidural opioids in children, as in adults, include excessive sedation, itching, nausea and vomiting, urinary retention, and delayed respiratory depression. Pruritus is often more of an annoyance than a major problem, and if it must be treated, one can use diphenhydramine (1.0 mg/kg) or low-dose naloxone (0.001–0.003 mg/kg per hour). Nausea and vomiting are particularly unpleasant side effects and may be treated with metoclopramide (0.1 mg/kg) or low-dose naloxone. Urinary retention may be a significant side effect, especially after genitourinary surgery if a urinary catheter is not left in place; in that situation, we usually choose an alternative method of pain relief. Although these side effects can be problems for some patients, the major risk of epidural opioids is delayed respiratory depression (50), which appears to be dose-related and which may require immediate treatment with oxygen, respiratory assistance, and naloxone (0.01 mg/kg).

Because of the possibility of delayed respiratory depression, we monitor all epidural opioid-treated patients with cardiorespiratory monitors for the period of risk, up to 12 hr after the last epidural injection. Moreover, we care for these patients in a setting (ward or intensive care unit) where the nursing staff has been thoroughly trained in the care of epidural catheters and in the recognition and treatment of all possible complications. To minimize the problems of delayed respiratory depression, we do not administer opioids by other routes during surgery or the postoperative period while epidural opioids are given to a patient.

**Regional Anesthesia.** Regional anesthesia with local anesthetic agents is useful in many situations for providing postoperative analgesia in children, and this technique has been employed extensively over the last few years. Here we briefly discuss only a couple of the most commonly used nerve blocks, and we refer the reader to recent reviews of the subject (51, 52).

The ilioinguinal-iliohypogastric nerve block with long-acting local anesthetics is effective after inguinal herniorrhaphy and orchiopexy. This block can be performed percutaneously by the surgeon or anesthesiologist prior to incision, or by the surgeon under direct visualization while the wound is open, with 0.25–0.5% bupivacaine (maximum dose 2.5 mg/kg). Simple instillation of local anesthetic in the open wound has been shown to be similarly effective (53). A “single shot” caudal injection of local anesthetic (e.g., 0.5–1.0 mL/kg bupivacaine, 0.125–0.25%, with epinephrine 1:200,000) may also be used for postoperative analgesia following herniorrhaphy and orchiopexy (as well as following lower-abdominal and lower-extremity surgery), but it may be associated with a higher incidence of urinary retention. For postoperative pain relief following circumcision, dorsal nerve block of the penis with 1–3 mL bupivacaine, 0.25% (epinephrine-free), is particularly effective. Finally, for a patient who has had lower-abdominal or lower-extremity surgery, continuous infusion of a local anesthetic agent (such as bupivacaine, 0.0625–0.125%) with or without opioid (for example, fentanyl) through an epidural catheter may provide excellent and safe postoperative analgesia. Such patients must be closely monitored.

## Summary

Postoperative pain management in children is a topic that has been neglected in the past but is currently an active field of interest and effort. Clearly, the child’s cognitive understanding of and emotional response to pain are different than an adult’s, and these differences make pain assessment and control more difficult. Ongoing work to develop more accurate techniques of estimating pain intensity in children may have helpful results. The effects of untreated pain in children are similar to those in adults but may have more long-term consequences in children. In the past, postoperative pain treatment in children was often inadequate, but newer techniques, such as continuous infusion of opioids, patient-controlled analgesia, epidural administration of opioids, and regional analgesia, hold promise for improved care in the future.

## References

1. National Center for Health Statistics, Division of Vital and Health Statistics, Publications Series 13, No. 82, 1983.
2. Unpublished data, National Center for Health Statistics, Division of Vital and Health Statistics, Bureau of Hospital Statistics, Graves E, personal communication, 1983.
3. Beyer JE, DeGood DE, Ashley LC, et al. Patterns of postoperative analgesic use with adults and children following surgery. Pain 1983; 17:71–81.
4. Mather L, Mackie J. The incidence of postoperative pain in children. Pain 1983; 15:271–282.
5. Schechter NL, Allen DA, Hanson K. Status of pediatric pain control: a comparison of hospital analgesic usage in children and adults. Pediatrics 1986; 77:11–15.
6. Anand KJS. The biology of pain perception in newborn infants. In: Tyler DC, Krane EJ, eds. Advances in pain research and therapy. Vol 15: Pediatric pain. New York: Raven Press, 1990:113–122.
7. Benedetti C, Bonica JJ, Bellucci G. Pathophysiology and therapy of postoperative pain: a review. Adv Pain Res Ther 1984; 7:373–407.
8. Beecher HK. Effect of laparotomy on lung volume. Demonstration of a new type of pulmonary collapse. J Clin Invest 1933; 12:651–658.
9. Beecher HK. The measured effect of laparotomy on the respiration. J Clin Invest 1933; 12:639–650.
10. Craig DB. Postoperative recovery of pulmonary function. Anesth Analg 1981; 60:46–52.
11. Mankikian B, Cantineau JP, Bertrand M, et al. Improvement of diaphragmatic function by a thoracic epidural block after upper abdominal surgery. Anesthesiology 1988; 68:379–386.
12. Wahba WM, Don HF, Craig DB. Postoperative epidural analgesia: effects on lung volumes. Can Anaesth Soc J 1975; 22:519–527.
13. Spence AA, Smith G. Postoperative analgesia and lung function: a comparison of morphine with extradural block. Br J Anaesth 1971; 43:144–148.
14. Nightingale DA, Richards CC. Volume-pressure relations of the respiratory system of curarized infants. Anesthesiology 1965; 26:710–714.
15. Richards CC, Bachman L. Lung and chest wall compliance of apneic paralyzed infants. J Clin Invest 1961; 40:273–278.
16. Mansell A, Bryan C, Levison H. Airway closure in children. J Appl Physiol 1972; 33:711–714.
17. Kavanagh C. A new approach to dressing change in the severely burned child and its effect on burn-related psychopathology. Heart Lung 1983; 12:612–619.
18. Yeager MP, Glass DD, Neff RK, et al. Epidural anesthesia and analgesia in high-risk surgical patients. Anesthesiology 1987; 66:729–736.
19. Abu-Saad H. Assessing children's responses to pain. Pain 1984; 19:163–171.
20. McGrath PA, deVeber LL, Hearn MT. Multidimensional pain assessment in children. Adv Pain Res Ther 1985; 9:387–393.
21. Scott PJ, Ansell BM, Huskisson EC. Measurement of pain in juvenile chronic polyarthritis. Ann Rheum Dis 1977; 36:186–187.
22. Beyer JE, Aradine CR. Content validity of an instrument to measure young children's perceptions of the intensity of their pain. J Pediatr Nurs 1986; 1:386–395.
23. McGrath PJ, Johnson G, Goodman JT, et al. CHEOPS: A behavioral scale for rating postoperative pain in children. Adv Pain Res Ther 1985; 9:395–402.
24. Krane EJ, Jacobson LE, Lynn AM, et al. Caudal morphine for postoperative analgesia in children: a comparison with caudal bupivacaine and intravenous morphine. Anesth Analg 1987; 66:647–653.
25. Gauvain-Piquard A, Rodary C, Rezvani A, et al. Pain in children aged 2–6 years: a new observational rating scale elaborated in a pediatric oncology unit—preliminary report. Pain 1987; 31:177–188.
26. Beyer JE, Knapp TR. Methodological issues in the measurement of children's pain. Child Health Care 1986; 14:233–241.
27. McGrath PA. An assessment of children's pain: a review of behavioral, physiological and direct scaling techniques. Pain 1987; 31:147–176.
28. McGrath PJ, Cunningham SJ, Goodman JT, et al. The clinical measurement of pain in children: a review. Clin J Pain 1986; 1:221–227.
29. McGrath PA. Pain assessment in children—a practical approach. In: Tyler DC, Krane EJ, eds. Advances in pain research therapy. Vol 15: Pediatric pain. New York: Raven Press, 1990:167–179.
30. Anand KJS, Hickey PR. Pain and its effects in the human neonate and fetus. N Engl J Med 1987; 317:1321–1329.
31. Chugani HT, Phelps ME. Maturational changes in cerebral function in infants determined by $^{18}$FDG positron emission tomography. Science 1986; 231:840–843.
32. Lipsitt LP. A study of sensory and learning processes of the newborn. Clin Perinatol 1977; 4:163–186.
33. Anand KJS, Sippell WG, Aynsley-Green A. Randomized trial of fentanyl anaesthesia in preterm babies undergoing surgery: effects on the stress response. Lancet 1987; 1:243–248.
34. Attia J, Amiel-Tison C, Mayer MN, et al. Measurement of postoperative narcotic administration in infants using a new clinical scoring system. Anesthesiology 1987; 67:A532.
35. Levine JD, Gordon NC. Pain in prelingual children and its evaluation by pain-induced vocalization. Pain 1982; 14:85–93.
36. Owens ME. Pain in infancy: conceptual and methodological issues. Pain 1984; 20:213–230.
37. Rigatto H, Brady JP, Verduzco R. Chemoreceptor reflexes in preterm infant. II. The effect of gestational and postnatal age on the ventilatory response to inhaled carbon dioxide. Pediatrics 1975; 55:614–620.
38. Way WL, Costley EC, Way EL. Respiratory sensitivity of the newborn infant to meperidine and morphine. Clin Pharm Ther 1965; 6:454–461.
39. Lynn AM, Slattery J. Morphine kinetics in early infancy. Anesthesiology 1987; 66:136–139.
40. Kupferberg HJ, Way EL. Pharmacological basis for the

increased sensitivity of the newborn rat to morphine. Pharmacol Exp Ther 1963; 141:105–112.

41. Lynn AM. Pharmacology of opioid compounds in pediatrics. In: Tyler DC, Krane EJ, eds. Advances in pain research and therapy. Vol. 15: Pediatric pain. New York: Raven Press, 1990:167–179.
42. Hertzka RE, Gauntlett IS, Fischer DM, Spellman MJ. Fentanyl-induced ventilatory depression: effects of age. Anesthesiology 1989; 70:213–218.
43. Berde CB, Sethna NF, Holzman RS, et al. Pharmacokinetics of methadone in children and adolescents in the postoperative period. Anesthesiology 1987; 67:A519.
44. Shanahan EC, Marshall AG, Garrett CPO. Adverse reactions to intravenous codeine phosphate in children. Anaesthesia 1983; 38:40–43.
45. Report of the committee on infectious diseases, American Academy of Pediatrics. Pediatrics 1982; 69:810–812.
46. Lynn AM, Opheim KE, Tyler DC. Morphine infusion after pediatric cardiac surgery. Crit Care Med 1984; 12:863–866.
47. Tyler DC. Patient controlled analgesia in adolescents. Pain 1987; 4:S236.
48. Jones SEF, Beasley JM, Macfarlane DWR, et al. Intrathecal morphine for postoperative pain relief in children. Br J Anaesth 1984; 56:137–141.
49. Krane EJ, Tyler DC, Jacobson LE. The dose response of caudal morphine in children. Anesthesiology 1989; 71:48–52.
50. Krane EJ. Delayed respiratory depression in a child after caudal epidural morphine. Anesth Analg 1988; 67:79–82.
51. Shandling B, Steward DJ. Regional analgesia for postoperative pain in pediatric outpatient surgery. J Pediatr Surg 1980; 15:477–480.
52. Yaster M, Maxwell LG: Pediatric regional anesthesia. Anesthesiology 1989; 70:324–338.
53. Casey WF, Rice LJ, Hannallah RS, et al. A comparison between bupivacaine instillation versus ilioinguinal/iliohypogastric nerve block for postoperative analgesia following inguinal herniorrhaphy in children. Anesthesiology 1990; 72:637–639.

# Acute Herpetic and Postherpetic Neuralgia: Clinical Features and Management

Bradley S. Galer, M.D., and Russell K. Portenoy, M.D.

## Abstract

Herpes zoster is a prevalent disorder that is transitory in most patients. Postherpetic neuralgia is defined by the persistence of pain following herpes zoster for some arbitrarily defined period, such as two months. Although most patients with postherpetic neuralgia experience a gradual remission of the pain, symptoms are prolonged in many, and intractable pain occurs in a few. This review describes the clinical characteristics and management of these syndromes.

Morbidity associated with the varicella-zoster virus is highly prevalent, especially among the elderly and immunocompromised. The recrudescence of this virus produces an acute segmental neuralgia known as acute herpes zoster or "shingles." Although the pain of the acute form of neuralgia may be severe, it is fortunately transient in most patients. Many patients, however, develop postherpetic neuralgia, a persistent pain that may become refractory to medical therapy. The high prevalence of acute and postherpetic neuralgia and the availability of potentially effective treatments for each underscore the importance of these syndromes and the need for clinicians to be aware of their clinical manifestations and management.

## Epidemiology

Age has a profound effect on the incidence and clinical characteristics of both forms of neuralgia (1–9). Children comprise 5%–8% of acute cases (3, 9), adults aged 50–70 years approximately 40% (3). Similarly, the overall incidence of acute neuralgia, which ranges from 1.3 to 4.8 cases per 1,000 person-years (3, 5, 8), is much lower than the incidence of 10.1 per 100 person-years in patients aged 80 years or older (3). The severity and duration of acute neuralgia (1, 9) and the incidence of postherpetic neuralgia also increase in the elderly (2, 10). In one study, the prevalence of postherpetic neuralgia one year after acute neuralgia was 4.2% for patients under 20 years of age and 47.5% for patients older than 70 years (2).

Like the elderly, immunocompromised patients also have an increased incidence of the acute condition (11–14). For example, cancer patients, especially those with hematologic or reticuloendothelial malignancies and those receiving immunosuppressive drugs, are at an increased risk for acute neuralgia (12, 14). No seasonal variation or sex or race preponderance has been documented in either form (1, 3, 5, 8–10).

## Pathology and Pathophysiology

**Acute Herpes Zoster.** The hypothesis that varicella zoster causes acute neuralgia was first proposed in 1943 (15) and has since withstood scientific inquiry. Acute varicella infection, usually a childhood disease, begins with viral entry

---

From the Department of Neurology, Albert Einstein College of Medicine, Bronx, NY (BSG), and the Pain Service, Department of Neurology, Memorial Sloan-Kettering Cancer Center, and Department of Neurology, Cornell University Medical College, New York, NY (RKP). Address reprint requests to Russell K. Portenoy, M.D., Director of Analgesic Studies, Pain Service, Dept. of Neurology, Memorial Sloan-Kettering Cancer Center, 1275 York Ave., New York, NY 10021.

through the oral or respiratory passages. Viremia then occurs with seeding of the skin. After the acute varicella infection, retrograde axonal transport of the virus to the dorsal root ganglion cells occurs and a latent phase ensues, usually persisting for many years. With reactivation, the virus spreads within the ganglion and returns to the skin by orthograde axonal transport. This reactivation is clinically manifested as acute neuralgia and is thought to occur when the host immune system is deficient, as in the immunocompromised patient (16). Immunologic deficiencies have been noted in the elderly, which helps to explain the increased incidence of this disorder among older patients (17).

In acute neuralgia, pathologic studies have shown that the affected dorsal root ganglia undergo acute inflammation and necrosis, often with hemorrhage (18–21). Although the virus has not been isolated from the ganglia (22, 23), viral particles and antigens have been identified in dorsal root ganglia, Schwann cells, peripheral nerves, and skin (24–27). Mild neuronal degeneration has also been seen in adjacent ganglia, as has degeneration of adjacent motor and sensory roots (18, 19), myelin breakdown, and axonal degeneration (25). Peripherally, a decrease in dermal nerve fibers surrounding the skin lesion has been observed (28, 29).

**Postherpetic Neuralgia.** The pathophysiologic changes that underlie the development of persistent pain following an incident of acute neuralgia are unknown. A postmortem analysis of one patient with thoracic postherpetic neuralgia revealed dorsal horn atrophy in several adjacent segments of the spinal cord on the affected side, with loss of myelin and axons (30). Also noted were fibrosis and cell loss in the ipsilateral dorsal root ganglion and large fiber loss in the ipsilateral dorsal root (30). Others have also observed preferential large nerve fiber loss in postherpetic neuralgia (31). In one case, immunohistochemical studies failed to demonstrate any changes in substance P, opioids, or serotonin (30).

Based on these findings, it has been suggested that pain in postherpetic neuralgia is due to the preferential loss of large inhibitory fibers, which results in diminished segmental inhibition (as modeled in the gate control theory of pain) and increased nociceptive transmission into the dorsal horn (30–32). Postherpetic neuralgia could also be due to ectopic activity in damaged or regenerating nociceptive afferent fibers (33), or perhaps activation of the nervi nervorum that invest nerve trunks (34).

Involvement of these peripheral mechanisms, although likely, cannot fully explain the syndrome. The long-term inefficacy of treatments that reduce or block peripheral input, such as cordotomy, suggests that a central mechanism must also contribute to the pain (7, 35). Many central mechanisms have been proposed (36), but none have yet been established in this condition.

A recent study suggests that there may be subgroups of postherpetic neuralgia patients with distinct pathophysiologic mechanisms (37). Two groups were distinguished according to characteristics of the pain, clinical examination, thermography, and response to local anesthetic treatment. One of these subgroups included those patients whose affected skin was extensively scarred, hypoesthetic, and cool, and who, on examination, revealed little or no allodynia (pain evoked by light touch). The second of these subgroups was composed of those patients who had minimally scarred, warm skin and who, on examination, proved to have near-normal sensation on the skin and severe allodynia (37).

Further investigation of such differences may suggest the use of specific therapeutic approaches to each subtype.

## Clinical Syndromes

**Acute Herpes Zoster.** The typical clinical appearance of acute herpes zoster is pain or itching in a dermatomal distribution, followed a few days later by a rash with the same distribution (20). Many patients, however, do not conform to this classic appearance. In one large series, over 20% of patients with this condition reported the onset of pain after the skin eruption, 7% developed pain and the rash simultaneously, and 0.5% had no pain (9). In addition, pain may precede the rash by as much as two weeks (20) or may occur without the development of any rash (zoster sine herpete) (38, 39).

The thoracic dermatomes are the most common site and are involved in more than 50% of cases (1, 5, 10, 11). The trigeminal nerve is involved in 3%–20%, the ophthalmic division being affected in the vast majority (1, 10, 40). The lumbar and cervical regions are involved in 10%–20%, and sacral involvement is least common. With increasing age, the proportion of patients with trigeminal acute herpes zoster increases, again usually within the ophthalmic division (1, 40). Bilateral zoster is known to occur, but in less than 1% of patients (1, 8, 20), and only 1%–8% of patients develop recurrences (1, 3, 6, 20), which occur at the site of prior eruption in half the cases (3, 6).

The pain of this condition is described as continuous, burning, or itching. Particularly when severe, the pain may be accompanied by marked psychological distress, disturbances in physical and social functioning, and anorexia and sleep disturbance.

**Postherpetic Neuralgia.** Postherpetic neuralgia is defined by the persistence of pain beyond healing of the acute condition. Because the natural history of acute herpetic neuralgia is gradual resolution of pain after vesicles heal, the point at which the acute condition is termed postherpetic is a matter of definition. A reasonable definition considers postherpetic neuralgia to be pain persisting for longer than two months after healing of the vesicles (7).

Overall, 12%–20% of patients with the acute condition complain of pain after vesicular healing (1, 6); four weeks after healing, pain persists in 9%, one-quarter of these patients having pain at one year (8). These rates are greatly affected by age, however, and in one large survey, pain persisted for more than one year in 50% of 70-year-old patients (2).

Postherpetic pain typically has several components, including constant burning, lacination, deep aching, and allodynia. Commonly associated with pain are depressed mood; functional disturbances; and vegetative signs, such as sleep disturbance, lassitude, anorexia, and weight loss (41).

## Treatment

### Acute Herpes Zoster

Numerous treatments have been reported to be effective in the acute condition, based on anecdotal observations. These include snake venom (42), parenteral epinephrine (43), the antifungal griseofulvin (44–46), the $H_2$-blocker cimetidine (47, 48), and the amebicides emetine (49) and dihydroemetine (50). The safety and efficacy of these treatments have not been verified in controlled trials, and none are currently recommended in the management of this condition.

Other treatments, including antivirals, corticosteroids, and anesthetics, have been scrutinized more thoroughly and have attained clinical acceptance.

**Antiviral Agents.** The current mainstay of treatment for acute herpes zoster is acyclovir. In a controlled study, oral acyclovir significantly increased the rate of rash resolution and reduced the pain (51). Prior to the availability of the oral formulation, two well-designed studies of parenteral acyclovir revealed similar results (52, 53).

Other antiviral compounds have also been investigated. A randomized, double-blind, placebo-controlled study of vidarabine demonstrated increased healing and a reduction in pain (54), and similar findings were obtained in studies using 5% idoxuridine ointment (55, 56) and interferon (57–59). Normal serum globulin, zoster immune globulin, and cystosine arabinoside have shown no effect in this condition (60, 61).

The most appropriate use of acyclovir during acute herpes zoster has not been fully defined. The drug is indicated only during the period of viral shedding, when lesions are visible. Immunocompromised patients at risk for viral dissemination should be treated promptly, usually parenterally. Many clinicians also use oral or parenteral acyclovir to treat selected immunocompetent patients, particularly the elderly with severe eruptions. Milder cases without severe skin damage or disabling pain do not appear to require treatment with this agent.

**Corticosteroids.** Corticosteroids are commonly used to treat the pain of acute herpes zoster. Early case reports and surveys observed benefits from both adrenocorticotropic hormone and steroids (62–66). A double-blind, placebo-controlled study using oral triamcinolone demonstrated that the drug improved pain but had no effect on healing time (67). More recently, a controlled study of prednisolone revealed a significant reduction in pain only in the first three days, corresponding to the period of vesicular inflammation, and no effect on rash resolution (68). Anecdotal reports of subcutaneous steroid injection have also suggested some benefit (69–72). Although viral dissemination has been reported following corticosteroid use in immunocompromised patients (20), no increased risk of dissemination in immunocompetent patients has been documented (7).

Based on these studies, immunocompetent patients with severe pain from acute herpes zoster are often given a trial of oral corticosteroids (for example, prednisone, 60–80 mg to start, with a gradual taper over two weeks). The response appears to be quite variable, but some patients demonstrate prompt reduction in pain and clearly benefit from the intervention.

**Anesthetic Approaches.** Many uncontrolled studies have reported relief of pain following regional anesthetic techniques in patients with the acute condition. Local anesthetics have been administered into the subcutaneous tissues (37, 73), peripheral nerves (73), epidural space (73–75), and paravertebral area (76). A mixture of an anesthetic and a corticosteroid has also been used, with favorable results (77, 78). These reports are all anecdotal, however, and neither subcutaneous infiltration nor somatic nerve blocks has achieved widespread acceptance.

Sympathetic blockade has also been advocated. In two uncontrolled studies, most patients obtained pain relief following one to four sympathetic blocks (79, 80), but timing of blockade seems crucial: pain relief occurred in 85%–90% of patients blocked within two weeks of onset, but in only 40% of those blocked after two weeks. Clinical experience with sympathetic blockade is extensive and, despite the absence of controlled studies, most practitioners believe that the technique can benefit some patients. A trial of sympathetic blockade is often suggested in immunocompromised patients who fail to benefit from acyclovir and in immunocompetent patients who are not relieved with steroids.

**Other Measures.** Care of the skin during the acute eruption, combined with the use of opioid and nonopioid analgesics, is also effective in ameliorating the pain of the acute condition. In milder cases, usually younger patients, the use of an opioid for one to two weeks is often adequate. Severely affected patients may benefit from administration of analgesic drugs concurrent with other, more specific approaches.

## Prevention of Postherpetic Neuralgia

Antiviral, corticosteroid, and anesthetic treatments have all been reported to reduce the incidence of postherpetic neuralgia if used during the acute condition. Unfortunately, this outcome has never been adequately verified in controlled studies.

Although a small controlled study claimed that postherpetic neuralgia was prevented if oral acyclovir treatment was initiated within 24 hr of onset of the acute disorder (81), a larger study failed to demonstrate any effect on the incidence or severity of postherpetic neuralgia at six months (82). Uncontrolled reports suggesting that corticosteroids reduce the incidence of postherpetic neuralgia (63, 65, 67, 83) were also not confirmed in a randomized, double-blind trial (68). Sympathetic blockade has been reported to reduce the incidence of the condition in some surveys (79, 80), but not others (84), and as noted, a controlled trial of this approach has not been performed.

There is, therefore, no effective means to prevent the development of persistent pain in patients with acute herpes zoster. A vaccine against varicella has been tested, and should this become available, primary prevention of this syndrome will be feasible. At the present time, however, treatment should not be implemented during the acute condition solely in the hope of preventing postherpetic neuralgia.

## Postherpetic Neuralgia

The literature is replete with claims of successful treatments for the postherpetic condition. Unfortunately, few approaches have been tested in controlled studies, and since most patients experience a spontaneous diminution of pain with time, conclusions drawn from small uncontrolled surveys are suspect. Adequate studies of analgesic efficacy must stratify the population by age, control for concurrent analgesic use, evaluate the type and degree of pain before and after treatment, and observe a lengthy follow-up interval (7). The need to define the type of pain has also been highlighted by the identification of postherpetic subgroups with potentially distinct responses to therapy (37).

**Systemic Pharmacologic Treatments.** Tricyclic antidepressants are widely considered a mainstay of treatment for established postherpetic neuralgia. Two controlled studies of amitriptyline have revealed significant improvement in pain without concurrent changes in depression ratings (85, 86). The median daily dose was 75 mg in one study (86) and 65 mg (range 12.5–150 mg) in the other (85). One of these studies (85) identified a positive correlation between analgesia and both amitriptyline dose and serum drug concentration (of both amitriptyline and its active metabolite, nortriptyline). These analgesic effects may result from activation of endogenous pain modulatory systems that employ serotonin or norepinephrine (85, 86).

Numerous surveys have reported improvement following treatment with a tricyclic combined with a neuroleptic (87–92) or an anticonvulsant (93–95). However, a placebo-controlled study in which a neuroleptic, chlorprothixene, was administered alone revealed only slight and transient benefit (96). Given the minimal support for efficacy and the potential for significant toxicity, especially in the elderly, neuroleptics should not be routinely combined with tricyclics, and trials of these drugs should be considered only in refractory cases.

A controlled study of carbamazepine revealed a significant reduction in lancinating postherpetic pain (97). This observation is consistent with the accepted view that anticonvulsants are useful in paroxysmal neuropathic pains, regardless of cause. In addition to carbamazepine, other anticonvulsants used for this indication include phenytoin, valproate, and clonazepam (7).

Lancinating pains that are refractory to anticonvulsant drugs may respond to other agents (98). Baclofen is a GABA agonist that has proven efficacy in trigeminal neuralgia (99) and is often used for stabbing neuropathic pains. The oral local anesthetic tocainide has also been demonstrated to be effective in trigeminal neuralgia (100), but a related compound, mexiletine, is preferred because of its greater safety. Mexiletine has also been shown to be efficacious in the treatment of continuous dysesthesias due to diabetic polyneuropathy (101); therefore, a trial of this drug can be recommended for continuous postherpetic pain refractory to tricyclic antidepressants. Finally, a recent controlled study revealed that the neuroleptic pimozide is as effective as carbamazepine in the treatment of trigeminal neuralgia (102). Consequently, pimozide can also be considered for refractory lancinating dysesthesias.

Although a study of single doses failed to reveal significant pain relief from either ibuprofen or codeine (103), some patients clearly benefit from the use of nonsteroidal drugs or opioids. Long-term opioid therapy for postherpetic pain, as for other nonmalignant pain syndromes, is controversial, but substantial data support this approach in selected patients (104).

**Topical Agents.** Topical treatments, including capsaicin, local anesthetic creams, and antiinflammatory ointments, have recently been introduced as potential therapies for postherpetic neuralgia. The advantages of topical agents include the reduced risk of systemic side effects and a decreased risk of drug interactions. Capsaicin depletes substance P and other peptide neurotransmitters from small primary afferent neurons (105, 106). In an open-label study using 0.025% capsaicin cream, pain decreased in 55% of patients who completed the trial (106). Another report suggested benefit in an even higher proportion of patients treated for six weeks with 0.75% capsaicin (105). Postapplication skin burning occurs in most patients treated with capsaicin and may preclude continued treatment in up to one-third (106). At present, the efficacy of this agent must be considered unproved, but anecdotal evidence is suggestive enough to recommend a trial of 0.025% capsaicin cream in patients who are refractory to other treatments or who would be placed at risk by other pharmacologic approaches. Clinical observation suggests that an adequate trial requires at least four weeks of treatment during which the cream is applied at least four times daily (105, 106).

Topical application of local anesthetics has also been observed to decrease postherpetic pain (107–109). Lidocaine ointment, both 5% and 10%, has had an analgesic effect in small nonblinded studies (107, 109). Of note, pain relief outlasted the duration of the anesthetic in most patients (107, 109). In one case, application of lidocaine-prilocaine cream provided pain relief lasting 11 hr (108). Topical local anesthetics may benefit only the constant burning component of the condition, not the paroxysmal lancinating component (107).

Treatment using local application of an antiinflammatory agent has also been reported. An aspirin-chloroform mixture was useful in one survey (110), but a controlled trial of benzydamine 3% cream failed to show any significant difference from placebo (111).

**Anesthetic Treatments.** Anesthetic procedures have a long history in the management of postherpetic neuralgia. Most reports, however, are anecdotal and lack adequate controls and follow-up. Sympathetic blockade provided only temporary relief in less than 50% of patients in several large series (79, 80, 112, 113). Poor results were reported following neurolytic blockade of peripheral nerves or roots (2, 114, 115). Intravenous administration of procaine yielded relief in less than one-half of patients (116), and intravenous lidocaine only transiently relieved paroxysmal pain in some patients (94). Transient benefit in a minority of patients was also observed following epidural or intrathecal injection of local anesthetic with or without steroid (77, 117). There is, therefore, little evidence to support the use of nerve blocks in the treatment of this condition. Mixed results have similarly been reported following local skin infiltration, with or without steroids (69, 71, 72, 118, 119).

Although the overall effects of anesthetic treatments have been disappointing, the observation that a few patients benefit from one or another of these approaches raises the interesting possibility that varying results relate to the existence of subgroups of postherpetic patients with different mechanisms of pain and, hence, different responses to treatment. One survey noted that only patients with allodynia obtained dramatic relief from direct injection of local anesthetic into the area of maximal pain (37). Studies are needed to evaluate the implications of this observation.

**Neurostimulatory Therapies.** Neurostimulatory modalities, such as counterirritation and transcutaneous electrical nerve stimulation (TENS), presumably activate endogenous pain-modulating systems, as modeled in the gate control theory (32). It can be conjectured that these methods would be particularly effective in diseases such as postherpetic neuralgia, in which a decreased number of large fibers has been documented (30, 31). Again, anecdotal clinical experience provides the bulk of the evidence for the efficacy of these modalities.

Counterirritation, or systematic rubbing of the affected area, has been reported to provide partial relief in some patients (120). This technique is often combined with the application of a vapocoolant spray. Application of the spray alone has also been reported to be effective (121).

Studies of TENS have reported mixed results. One survey observed moderate to complete pain relief outlasting stimulation in approximately one-half of patients (122), and others have reported similar benefit (87, 123–125). One study revealed only transient relief (126), however, and another claimed no benefit from TENS (93). Given the safety of the approach, a trial of TENS is reasonable in all patients with refractory postherpetic neuralgia.

Other, more invasive forms of neurostimulation have been attempted. Acupuncture yielded no benefit in a randomized, single-blind study (127). Dorsal column stimulation, used in a variety of chronic pain syndromes (128, 129), has not been thoroughly tested for postherpetic neuralgia; experience with this technique suggests that sustained relief is uncommon (130). Small series of patients treated with deep brain stimulation have also been described (131–133); thalamic stimulation produced good to excellent relief in approximately 50% of patients with long-standing trigeminal postherpetic neuralgia (131, 132).

Given the minimal evidence for efficacy and the potential risks involved, invasive techniques of neurostimulation should be considered only in patients with a disabling condition refractory to other approaches. These techniques should be implemented only by experienced practitioners.

**Surgical Procedures.** With the exception of dorsal root entry zone (DREZ) lesion, surgical approaches have yielded disappointing results in patients with postherpetic neuralgia. Neurectomy, rhizotomy, sympathectomy, and cordotomy have provided only transient benefit (2, 115, 125, 134–139), although rare cases of sustained relief have been reported (134, 139). Neuroablation of brainstem nuclei or tracts, mesencephalotomy or mesencephalothalamotomy (140), and ventrobasal or medial thalamotomy (141) have also been performed but have provided relief only in isolated cases.

Results in refractory postherpetic neuralgia appear to be more favorable with the DREZ lesion. One series reported immediate relief in all 17 patients undergoing DREZ lesioning and sustained relief in 14 of 17 (142). Further studies of this technique as treatment for postherpetic neuralgia are needed, but the approach has gained some clinical acceptance for patients with disabling thoracic postherpetic neuralgia unresponsive to other measures.

Surgical undermining or resection of scarred skin has also been used rarely in this condition. These approaches have reportedy yielded improvement in less than one-third of patients (143, 144), results similar to other, safer treatments aimed at the skin, such as injection of local anesthetics or steroids and cryoanalgesia (7).

## Suggested Management

**Acute Herpes Zoster.** All patients with acute herpes zoster should have appropriate skin care and access to analgesic drugs, including opioids. Acyclovir is recommended for all immunocompromised patients and for those immunocompetent patients with severe pain or skin damage. The latter can usually be treated orally. Based on a controlled trial, a dose of 800 mg five times daily can be administered orally for seven days (51). Acyclovir therapy is generally appropriate only when lesions are still present.

Although corticosteroids are commonly used, their efficacy remains controversial (7, 145). Based on a recent well-controlled study, steroid therapy should be considered soon after the onset of the acute condition in the immunocompetent patient with severe pain (68). As noted, an acceptable regimen is 60–80 mg of prednisone daily, with a taper period of 7–10 days (7).

Although never subjected to rigorous evaluation, sympathetic blockade appears to be useful in some patients with the acute condition. Patients who continue to experience disabling pain despite acyclovir, analgesic drugs, and possibly corticosteroid therapy should be considered for sympathetic blockade. Results can be anticipated to be much better if the procedure is performed within the first two weeks of eruption (79, 80).

If the patient is refractory to these therapies, consideration should be given to the use of treatments typically initiated for postherpetic neuralgia, including a tricyclic antidepressant, anticonvulsant, or other drugs as indicated. After healing of the skin, a trial of TENS can be implemented.

**Postherpetic Neuralgia.** Although there is no thoroughly reliable treatment for postherpetic neuralgia, many therapies have been shown to be efficacious, at least for some patients with this disorder. Treatment depends on comprehensive assessment, which should include an accurate description of the patient's pain (constant burning, intermittent shooting, allodynia) and an evaluation of the patient's psychological status, level of physical functioning, concurrent medical problems, and premorbid condition. The need for a multimodal, multidisciplinary approach to therapy should be assessed in all patients. Many patients benefit from concurrent administration of several therapies, including one or more drugs, TENS, physical therapy, and psychological interventions designed to augment pain relief and enhance functional capabilities. At the end of the clinical spectrum, patients with severe pain and disability may be appropriate candidates for referral to a multidisciplinary pain management program, which provides an integrated multimodal approach with a strong rehabilitative focus. Less severe cases are appropriately managed by individual clinicians, who must carefully assess the patient's requirements, refer as needed for physical therapy and psychological treatment, and implement an analgesic regimen.

One approach to primary analgesic therapies (Fig.) is empirical and based on studies of patients with postherpetic neuralgia and experience in other neuropathic pain syndromes. The efficacy of all treatments depends on optimal administration. For drug therapies, this usually requires a trial of at least several weeks during which low initial doses are gradually increased to a therapeutic range; for example, amitriptyline is typically started at a dose of 10–25 mg at night, but most patients require doses greater than 50 mg to demonstrate any response, and some patients require doses greater than 150 mg daily. Drug combinations are commonly used, and careful monitoring is required to limit the risks of this approach and ensure that each agent provides independent, and therefore additive, relief. Throughout this process, the dual goals of treatment—pain relief and improved function—must be continually emphasized.

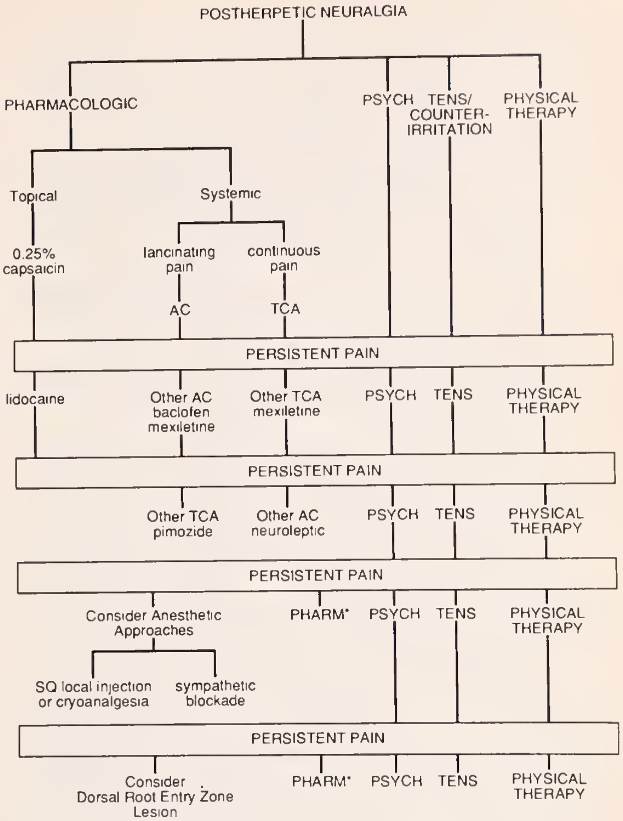

POSTHERPETIC NEURALGIA
PHARMACOLOGIC
PSYCH
TENS/ COUNTER-IRRITATION
PHYSICAL THERAPY
Topical
Systemic
lancinating pain
continuous pain
0.25% capsaicin
AC
TCA
PERSISTENT PAIN
lidocaine
Other AC baclofen mexiletine
Other TCA mexiletine
PSYCH
TENS
PHYSICAL THERAPY
PERSISTENT PAIN
Other TCA pimozide
Other AC neuroleptic
PSYCH
TENS
PHYSICAL THERAPY
PERSISTENT PAIN
Consider Anesthetic Approaches
PHARM*
PSYCH
TENS
PHYSICAL THERAPY
SQ local injection or cryoanalgesia
sympathetic blockade
PERSISTENT PAIN
Consider Dorsal Root Entry Zone Lesion
PHARM*
PSYCH
TENS
PHYSICAL THERAPY

Fig. Algorithm for management of established postherpetic neuralgia. Following comprehensive assessment, consideration should be given to use of multimodal approach, beginning with use of pharmacologic, neurostimulatory, psychological, and physiatric techniques. Persistent pain requires sequential drug trials and additional emphasis on psychological and physiatric interventions. Severe refractory pain may ultimately lead to consideration of invasive approaches. Management of patients with difficult disorders may be clarified through evaluation by specialists in a comprehensive multidisciplinary pain management setting. AC, anticonvulsant; TCA, tricyclic antidepressant; PHARM*, any pharmacologic agent that has proved even minimally efficacious should be continued.

## References

1. Burgoon CF, Burgoon JS, Baldridge GD. The natural history of herpes zoster. JAMA 1957; 164:265–269.
2. Demorgas JM, Kierland RR. The outcome of patients with herpes zoster. Arch Dermatol 1957; 75:193–196.
3. Hope-Simpson RE. The nature of herpes zoster: a long-term study and a new hypothesis. Proc R Soc Lond (Biol) 1965; 58:9–20.
4. Hope-Simpson RE. Postherpetic neuralgia. J R Coll Gen Pract (Occas Pap) 1975; 25:571–575.
5. McGregor RM. Herpes zoster, chicken pox and cancer in general practice. Br Med J 1957; 1:84–87.
6. Molin GJ. Aspects of the natural history of herpes zoster. Acta Dermatol Venereol (Stockh) 1969; 49:569–572.
7. Portenoy RK, Duma C, Foley KM. Acute herpetic and postherpetic neuralgia: clinical review and current management. Ann Neurol 1986; 20:651–664.
8. Ragozzino MW, Melton LJ, Kurland LT, et al. Population based study of herpes zoster and its sequelae. Medicine (Baltimore) 1982; 61:310–316.

9. Rogers RS, Tindall JP. Geriatric herpes zoster. J Am Geriatr Soc 1971; 19:495–503.
10. Watson CPN, Evans RJ, Watt VR, Birkett N. Postherpetic neuralgia: 208 cases. Pain 1988; 35:289–297.
11. Mazur M, Dolin R. Herpes zoster at the NIH: a twenty year history. Am J Med 1978; 65:738–744.
12. Reboul F, Donaldson SS, Kaplan HS. Herpes zoster and varicella infection in children with Hodgkin's disease: an analysis of contributing factors. Cancer 1978; 41:95–99.
13. Rifkind D. The activation of varicella-zoster virus infection by immunosuppressive therapy. J Lab Clin Med 1966; 68:463–474.
14. Schimpff S, Serpick MD, Stoller B, et al. Varicella-zoster infections in patients with cancer. Ann Intern Med 1972; 76:241–254.
15. Garland J. Varicella following exposure to herpes zoster. N Engl J Med 1943; 228:336–337.
16. Price RW. Neurobiology of human herpes virus infections. In: Roses AD, ed. CRC critical reviews in clinical neurobiology. Vol 2. Boca Raton, FL: CRC Press, 1985:61–123.
17. Miller AE. Selective decline in cellular immune response to varicella-zoster in the elderly. Neurology 1980; 30:582–587.
18. Denny-Brown D, Adams RD, Fitzgerald PJ. Pathologic features of herpes zoster: a note on "geniculate" herpes. Arch Neurol Psychiatry 1944; 51:216–231.
19. Head H, Campbell AW. The pathology of herpes zoster and its bearing on sensory localization. Brain 1900; 23:353–523.
20. Juel-Jensen BE, MacCallum FO. Herpes simplex, varicella and zoster. Philadelphia: Lippincott, 1972:127–132.
21. Smith FP. Pathological studies of spinal nerve ganglia in relation to intractable intercostal pain. Surg Neurol 1978; 10:50–53.
22. Tenser RB. Herpes simplex and herpes zoster: nervous system involvement. Neurol Clin 1984; 2:229–234.
23. Weiler TH. Varicella and herpes zoster. N Engl J Med 1982; 309:1362–1368, 1434–1439.
24. Bastian FO, Rebson AS, Yee CL, Troika TS. Herpes viral varicellae isolated from human dorsal root ganglia. Arch Pathol 1974; 97:331–333.
25. Esiri MM, Tomlinson AH. Herpes zoster: demonstration of virus in trigeminal nerve and ganglion by immunofluorescence and electron microscopy. J Neurol Sci 1972; 15:35–48.
26. Gharak NR, Zimmerman HM. Spinal ganglion in herpes zoster. Arch Pathol 1973; 95:411–415.
27. Nagashima K, Nakazawa M, Endo H, et al. Pathology of the human spinal ganglia in varicella-zoster virus infection. Acta Neuropathol (Berl) 1975; 33:105–117.
28. Ebert MH. Histologic changes in sensory nerves of the skin in herpes zoster. Arch Dermatol 1949; 60:641–648.
29. Muller SA, Winkelmann RK. Cutaneous nerve changes in zoster. J Invest Dermatol 1969; 52:71–77.
30. Watson CPN, Morstead C, Van der Kaug D, Evans RJ. Post-herpetic neuralgia: post-mortem analysis of a case. Pain 1988; 34:129–138.
31. Noordenbos W. Physiological correlates of clinical pain syndromes. In: Soulairac A, Cahn J, Charpentier J, eds. Pain. New York: Academic Press, 1968:465–476.
32. Melzack R, Wall PD. Pain mechanisms: a new theory. Science 1965; 150:971–979.
33. Asbury AK, Fields HL. Pain due to peripheral nerve damage: a hypothesis. Neurology 1984; 34:1587–1590.
34. Scadding JW. Peripheral neuropathies. In: Wall PD, Melzack R, eds. Textbook of pain. New York: Churchill Livingstone, 1989:522.
35. Wall PD. Alteration in the central nervous system after deafferentation connectivity control. In: Bonica JJ, Lindblom U, Iggo A, eds. Advances in pain research and therapy, vol. 5. New York: Raven, 1983:677–690.
36. Portenoy RK: Issues in the management of neuropathic pain. In: Basbaum A, Besson J-M, eds. Towards a new pharmacotherapy of pain. New York: Wiley, 1991, 393–416.
37. Rowbotham MC, Fields HL. Post-herpetic neuralgia: the relation of pain complaint, sensory disturbance, and skin temperature. Pain 1989; 39:129–144.
38. Easton HG. Zoster sine herpete causing acute trigeminal neuralgia. Lancet 1970; 2:1065.
39. Lewis GW. Zoster sine herpete. Br Med J 1958; 2:418–421.
40. Brown GR. Herpes zoster: correlation of age, sex, distribution, neuralgia and associated disorders. South Med J 1976; 69:576–578.
41. Sternbach R. Pain patients: traits and treatment. New York: Academic Press, 1974.
42. Dennie CC, Morgan DB, Combs FP. Treatment of herpes zoster with moccasin venom. J Invest Dermatol 1949; 12:153–155.
43. Duke WW. Epinephrine in the treatment of the pain of herpes zoster. JAMA 1924; 83:1919.
44. Chaturvedi U, Mathur A. Griseofulvin in the treatment of herpes zoster. Br Med J 1975; 3:438.
45. Joelson L. Griseofulvin in the tretment of herpes zoster. S Afr Med J 1978; 54:638.
46. Joubert J. Griseofulvin in the treatment of herpes zoster. S Afr Med J 1978; 54:224.
47. Maulight GM, Talpaz M. Cimetidine for herpes zoster. N Engl J Med 1984; 310:318–319.
48. VanderSpuy S, Levy DW, Levin W. Cimetidine in the treatment of herpesvirus infection. S Afr Med J 1980; 58:112–116.
49. Hoehn G. Emetine in herpes zoster. Cutis 1981; 27:10.
50. Hernandez-Perez R. Dihydroemetine therapy for herpes zoster: a comparison with corticosteroids. Cutis 1980; 25:424–426.
51. McKendrick MW, McGill JI, White JE, Wood MJ. Oral acyclovir in acute herpes zoster. Br Med J 1986; 293:1529–1532.
52. Bean B, Braun C, Balfour H. Acylovir therapy for acute herpes zoster. Lancet 1982; 2:118–121.
53. Peterslund NA, Ipsen J, Schonheyden H, et al. Acyclovir in herpes zoster. Lancet 1981; 2:827–830.
54. Whitley RJ, Soong SJ, Dolin R, et al. Early vidarabine therapy to control the complications of herpes zoster in immune suppressed patients. N Engl J Med 1982; 307:971–975.
55. Dowber R. Idoxuridine in herpes zoster: further evaluation of intermittent topical therapy. Br Med J 1974; 2:526–527.
56. Juel-Jensen BE, MacCallum FO, McKenzie A, et al. Treatment of zoster with idoxuridine in dimethyl sulphoxide: results of two double-blind controlled trials. Br Med J 1970; 4:776–780.
57. Emodi G, Rufli T, Just M, et al. Human interferon therapy for herpes zoster in adults. Scand J Infect Dis 1975; 7:1–5.
58. Merigan T, Rand K, Pollard P, et al. Human leukocyte

interferon for the treatment of herpes zoster in patients with cancer. N Engl J Med 1978; 298:981–987.
59. Usuki K, Kitamura K, Urabe A, Takaku F. Herpes zoster, postherpetic neuralgia, and interferon-gamma. Ann Intern Med 1988; 108:643–644.
60. Betts RF, Zaky DA, Douglas RG, Roger G. Ineffectiveness of subcutaneous cytosine arabinoside in localized herpes zoster. Ann Intern Med 1975; 82:778–783.
61. Stevens DA, Merigan TC. Zoster immune globulin prophylaxis of disseminated zoster in compromised hosts. Arch Intern Med 1980; 140:52–54.
62. Appelman DH. Treatment of herpes zoster with ACTH. N Engl J Med 1955; 253:693–695.
63. Elliot FA. Treatment of herpes zoster with high doses of prednisone. Lancet 1964; 2:610–611.
64. Frank L, Lysiak R. Herpetic and postherpetic pain treated with cortisone and ACTH. NY State Med J 1953; 53:2379.
65. Keczkes K, Basheer AM. Do corticosteroids prevent postherpetic neuralgia? Br J Dermatol 1980; 102:551–555.
66. Pearse J. Postherpetic neuralgia. Br Med J 1973; 1:679.
67. Eaglstein WH, Katz R, Brown JA. The effects of early cortiocosteroid therapy on the skin eruption and pain of herpes zoster. JAMA 1970; 211:1681–1683.
68. Esmann V, Kroor S, Peterslund NA, et al. Prednisolone does not prevent post-herpetic neuralgia. Lancet 1987; 8551:126–129.
69. Epstein E. Treatment of herpes zoster and postzoster neuralgia by subcutaneous injection of triamcinolone. Int J Dermatol 1981; 20:65–68.
70. Epstein E. Treatment of herpes zoster and postzoster neuralgia by the sublesional injection of triamcinolone and procaine. Acta Derm Venereol (Stockh) 1970; 50:69–73.
71. Epstein E. Treatment of zoster and postzoster neuralgia by the intralesional injection of triamcinolone: a computer analysis of 199 cases. Int J Dermatol 1976; 15:762–769.
72. Epstein E. Triamcinolone-procaine in the treatment of zoster and postzoster neuralgia. Calif Med 1971; 115:6–10.
73. Riopelle JM, Naraghi M, Grush KP. Chronic neuralgia incidence following local anesthetic therapy for herpes zoster. Arch Dermatol 1984; 120:747–750.
74. Bauman J. Treatment of acute herpes zoster neuralgia by epidural injection of stellate ganglion block. Anesthesiology 1979; 51:223.
75. Marmer MJ. Acute herpes zoster: successful treatment by continuous epidural analgesia. Calif Med 1965; 103:277–279.
76. Johnson LR, Rocco AG, Ferrante FM. Cutaneous subpleural-paravertebral block in acute thoracic herpes zoster. Anesth Analg 1988; 67:1105–1108.
77. Perkins HM, Hanlon PR. Epidural injection of local anesthetic and steroids for relief of pain secondary to herpes zoster. Arch Surg 1978; 113:253–254.
78. Schreuder M. Pain relief in herpes zoster. S Afr Med J 1983; 63:820–821.
79. Colding A. The effect of sympathetic blocks on herpes zoster. Acta Anaesthesiol Scand 1969; 13:133–141.
80. Colding A. Treatment of pain: organization of a pain clinic: treatment of acute herpes zoster. Proc R Soc Lond (Biol) 1973; 66:541–543.
81. Finn R, Smith MA. Oral acyclovir for herpes zoster. Lancet 1984; 2:575.
82. McKendrick MW, McGill JF, Wood MJ. Lack of effect of acyclovir on postherpetic neuralgia. Br Med J 1989; 298:431.
83. Finn R. Prednisolong does not prevent post-herpetic neuralgia (Letter). Lancet 1987; 8558:576.
84. Yahagida H, Suwa K, Corsenn G. No prophylactic effect of early sympathetic blockade on postherpetic neuralgia. Anesthesiology 1987; 66:73–76.
85. Max MB, Schafer SC, Culnane M, Smoller B, Dubner R, Gracely RH. Amitriptyline, but not lorazepam, relieves postherpetic neuralgia. Neurology 1988; 38:1427–1432.
86. Walsh TD. Antidepressants in chronic pain. Clin Neuropharmacol 1983; 6:271–295.
87. Crue BL, Todd EM, Moline DB. Postherpetic neuralgia—conservative treatment regimen. In: Crue BL, ed. Pain: research and treatment. New York: Academic Press, 1975:289–292.
88. Kocher R. Use of psychotropic drugs for the treatment of chronic severe pain. In: Bonica JJ, Albe-Fessard D, eds. Advances in pain research and therapy. Vol. 1. New York: Raven Press, 1976:579–582.
89. Merskey H, Hester RA. Treatment of chronic pain with psychotropic drugs. Postgrad Med J 1972; 48:594–598.
90. Taub A. Relief of postherpetic neuralgia with psychotropic drugs. J Neurosurg 1973; 39:235–239.
91. Taub A, Collins WF. Observations on the treatment of dysesthesia with psychotropic drugs. Adv Neurol 1974; 4:309–315.
92. Weis O, Sriwaranakul K, Weintraub M. Treatment of postherpetic neuralgia and acute herpetic pain with amitriptyline and perphenazine. South Med J 1982; 62:274–275.
93. Gerson GR, Jones RB, Luscombe DK. Studies on the concomitant use of carbamazepine and clomipramine for the relief of post-herpetic neuralgia. Postgrad Med J 1977; 54:104–109.
94. Hatangdi VS, Boas RA, Richards EG. Postherpetic neuralgia: management with antiepileptic and tricyclic drugs. In: Bonica JJ, Albe-Fessard D, eds. Advances in pain research and therapy. Vol. 1. New York: Raven Press, 1976:583–587.
95. Raftery H. The management of postherpetic pain using sodium valproate and amitriptyline. Irish Med J 1979; 72:399–401.
96. Nathan PW. Chlorprothixene (Taractan) in postherpetic neuralgia and other severe chronic pains. Pain 1978; 5:367–371.
97. Killian JM, Fromm GH. Carbamazepine with treatment of neuralgia. Arch Neurol 1968; 19:129–136.
98. Portenoy RK. Pharmacologic management of chronic pain. In: Fields HL, ed. Pain syndromes in neurology. London: Butterworth, 1990:257–278.
99. Fromm GH, Terence CF, Chatta AS. Baclofen in the treatment of trigeminal neuralgia. Ann Neurol 1984; 15:240–247.
100. Lindstrom P, Lindblom U. The analgesic effect of tocainide in trigeminal neuralgia. Pain 1987; 28:45–50.
101. Dejgard A, Petersen P, Kastrup J. Mexiletine for the treatment of chronic painful diabetic neuropathy. Lancet 1988; 1:9–11.
102. Lechin F, et al. Pimozide therapy for trigeminal neuralgia. Arch Neurol 1989; 46:960–964.
103. Max MB, Schafer SC, Culnane M, Dubner R, Gracely RH. Association of pain relief with drug side effects in postherpetic neuralgia: a single-dose study of clonidine, codeine, ibuprofen and placebo. Clin Pharmacol Ther 1987; 43:363–371.

104. Portenoy RK, Foley KM. Chronic use of opioid analgesics in nonmalignant pain: report of 38 cases. Pain 1986; 25:171–186.
105. Bernstein JE, et al. Topical capsaicin treatment of chronic postherpetic neuralgia. J Am Acad Dermatol 1989; 21:265–270.
106. Watson CPN, Evans RJ, Watt UR. Postherpetic neuralgia and topical capsaicin. Pain 1988; 33:333–340.
107. Kissin I, McDonald J, Xavier AV. Topical lidocaine for relief of superficial pain in postherpetic neuralgia. Neurology 1989; 39:1132–1133.
108. Milligan KA, Atkinson RE, Schofield PA. Lignocaine-prilocaine cream in postherpetic neuralgia. Br Med J 1989; 298:253.
109. Rowbotham MC, Fields HL. Topical lidocaine reduces pain in post-herpetic neuralgia. Pain 1989; 38:297–301.
110. King RB. Concerning the management of pain associated with herpes zoster and of postherpetic neuralgia. Pain 1988; 33:73–78.
111. McQuay HJ, Caroll D, Moxon A, Glynn CJ, Moore RA. Benzydamine cream for the treatment of postherpetic neuralgia. Pain 1990; 40:131–135.
112. Masud KZ, Forster KJ. Sympathetic block in herpes. Am Fam Physician 1975; 12:142–143.
113. Milligan NS, Nash TP. Treatment of post-herpetic neuralgia: a review of 77 consecutive cases. Pain 1985; 23:381–386.
114. Peer MM. Postherpetic trigeminal neuralgia: persistence of pain after section of the sensory root of the gasserian ganglion. JAMA 1929; 92:1503–1505.
115. Sugar O, Bucy PC. Post-herpetic trigeminal neuralgia. Arch Neurol Psychiatry 1951; 65:131–145.
116. Shanbrom E. Treatment of herpetic pain and postherpetic neuralgia with intravenous procaine. JAMA 1961; 12:1041–1043.
117. Sehgal AD, Gardner WJ. Postherpetic neuralgia. Neurology 1962; 12:725–727.
118. Lefkovitz AM. Postherpetic neuralgia. Neurology 1961; 1:1.
119. Russell WR, Espir MLE, Morganstern FS. Treatment of post-herpetic neuralgia. Lancet 1957; 1:242–245.
120. Todd EM, Crue BL, Vergadamo M. Conservative treatment of post-herpetic neuralgia. Bull Los Angeles Neurol Soc 1965; 30:148.
121. Taverner D. Alleviation of post-herpetic neuralgia. Lancet 1960; 2:671–673.
122. Nathan PW, Wall PD. Treatment of post-herpetic neuralgia by prolonged electrical stimulation. Br Med J 1974; 3:645–647.
123. Haas LF. Postherpetic neuralgia. Trans Ophthalmol Soc NZ 1977; 29:133–136.
124. Long DM. External electrical stimulation as a treatment of chronic pain. Minn Med 1974; 57:195–198.
125. Long DM. Pain of peripheral nerve injury. In: Youmans JR, ed. Neurological surgery. Philadelphia: WB Saunders, 1982:3641–3643.
126. Loeser JD, Black RG, Christman A. Relief of pain by transcutaneous stimulation. J Neurosurg 1975; 42:308–314.
127. Lewith GT, Field J, Machin D. Acupuncture compared with placebo in postherpetic pain. Pain 1983; 17:361–368.
128. Burton C. Dorsal column stimulation: optimization of application. Surg Neurol 1975; 4:171–176.
129. Nielson KD, Adams JE, Hosobuchi Y. Experience with dorsal column stimulation for relief of chronic intractable pain: 1968–1973. Surg Neurol 1975; 4:148–152.
130. Hitchcock E. A comparison of analgesic ablative and stimulation techniques. Zentralbl Neurochir 1981; 42:189–199.
131. Mazars GJ, Merienne L, Cioloca C. Comparative study of electrical stimulation of posterior thalamic nuclei, periaqueductal gray and other midline mesencephalic structures in man. In: Bonica JJ, Liebeskind JC, Albe-Fessard D, eds. Advances in pain research and therapy. Vol. 3. New York: Raven Press, 1979:541–546.
132. Siegfried J. Monopolar electrical stimulation of ventroposteriomedialis thalami for postherpetic facial pain. Appl Neurophysiol 1982; 45:179–184.
133. Young RF, Kroening R, Fulton W, et al. Electrical stimulation of the brain in treatment of chronic pain. J Neurosurg 1985; 62:389–396.
134. Cushing H. The major trigeminal neuralgias of their surgical treatment based on experience with 332 gasserian operations. Am J Med Sci 1920; 160:157–184.
135. Hyndman OR. Postherpetic neuralgia in the distribution of cranial nerves: evidence for sympathetic mediation and surgical cure. Arch Neurol Psychiatry 1939; 42:222–232.
136. Lipton S. Percutaneous cordotomy. In: Wall PD, Melzack R, eds. Textbook of pain. New York: Churchill Livingstone, 1984:632–638.
137. Onofrio BM, Campa HK. Evaluation of rhizotomy. Review of 12 years experience. J Neurosurg 1972; 36: 751–755.
138. Voris HC, Whistler WW. Treatment of Pain. Springfield, IL: Thomas, 1973:82.
139. White JC, Sweet WH. Pain and the neurosurgeon: a forty year experience. Springfield, IL: Thomas, 1969:472–477.
140. Wycis HT, Spiegel EH. Long-range results in the treatment of intractable pain by stereotactic midbrain surgery. J Neurosurg 1962; 19:101–107.
141. Hitchcock ER, Teixeira MJ. A comparison of results from center-median and basal thalamotomies for pain. Surg Neurol 1981; 15:341–351.
142. Friedman AH, Nashold BS. DREZ lesions for postherpetic neuralgia. Neurosurgery 1984; 15:969–970.
143. Tindall GT, Odom GL, Vieth RG. Surgical treatment of postherpetic neuralgia: results of skin undermining and excision in 14 patients. Arch Neurol 1962; 7:423–426.
144. Weidman MJ. Surgical relief of post-herpetic neuralgia by excision of scarred skin. Med J Aust 1976; 63:472–473.
145. Post BT, Philbrook JT. Do corticosteroids prevent postherpetic neuralgia? A review of the evidence. J Am Acad Dermatol 1988; 18:605–610.

# Organization of a Multidisciplinary Pain Center

JOHN C. ROWLINGSON, M.D., AND ROBIN J. HAMILL, M.D.

PAIN management centers embody the fundamental concept that patients with complex pain problems are best served by a team of specialists with different health care backgrounds. This team facilitates the comprehensive evaluation of patients with pain and provides reputable, contemporary pain management (1–6). All practitioners involved in the care of these patients share the data gathered, including information on physical, biochemical, ethnocultural, cognitive, affective, environmental, and psychosocial factors. New understanding of the nature of pain and pain-related behavior is being developed and applied (4, 6) in pain management centers and pain clinics so that more accurate diagnosis and appropriate therapy can follow.

Many different practice styles are used in achieving the coordination of services and specialists that make up a pain center (2, 4–6). Solo practitioners may have only a referral network in their community. Larger practices may permit three or more specialists to work in the same building. Pain clinics associated with teaching hospitals are likely to have extensive programs incorporating physicians and pain management specialists of virtually every possible interest, as well as inpatient services (1–4, 6).

Inpatient services are expensive in both time and resources and can serve only a limited number of patients. However, hospitalization is an excellent option for patients requiring drug detoxification, extensive psychological or behavioral therapy, guided physical therapy, or nonambulatory care. In addition, inpatient therapy may benefit those patients for whom outpatient therapy has failed or for whom life disruption is global, and who thus need temporary removal from a detrimental home situation to refocus their lives away from the pain (1, 4, 5).

Outpatient therapy costs far less and puts the burden for improvement squarely on patients' shoulders because they must adhere to treatment plans amid the distractions in their home environment. Also, activities of daily living, such as driving, can be incorporated into the therapy program and utilized while the patient continues to work. Finally, outpatient management demands that the local or referring physicians be educated about the patient's therapy and also be actively involved in implementation of the treatment plan so that a consistent ideologic approach is provided away from the pain center.

Chronic pain problems can stymie and frustrate referring physicians to the point of ineffectiveness. A comprehensive, second-opinion approach to referred patients is necessary because the patient who has been disappointed by uncertain diagnoses and unsuccessful treatment can also suffer anxiety, depression, and desperation. The pain center staff has concentrated experience with complex pain problems. Staff members apply established protocols that are continually upgraded and adjusted to meet the needs of individual patients. The abundant and broad expertise available at a multidisciplinary center encourages consideration of emotional, psychosocial, familial, and occupational consequences of the patient's pain, critical factors that can be overlooked in the chain of single-subspecialty evaluations. The comprehensive body of data obtained paves the way for development of a treatment plan that addresses all aspects of the "pain."

---

From the Pain Management Center, Department of Anesthesiology, University of Virginia Health Sciences Center, Charlottesville, Virginia.

Address reprint requests to John C. Rowlingson, M.D., Professor of Anesthesiology and Director, Pain Management Center, Department of Anesthesiology, Box 238, University of Virginia Health Sciences Center, Charlottesville, VA 22908.

Attempts at establishing standardization of facilities and services are a natural part of the evolution of pain clinics. The leader in the field has been the Commission for Accreditation of Rehabilitation Facilities. As one might anticipate, there is an expectation for any pain clinic so accredited to match the standards of a rehabilitation facility, but every pain clinic and pain management center is not necessarily a rehabilitation center. Therefore, other organizations are becoming involved. The International Association for the Study of Pain recently generated a report that provides standards for interdisciplinary pain treatment units run by those of many specialties—psychology, psychiatry, neurosurgery, anesthesiology, physical and rehabilitation medicine—all prominent in IASP membership. The American Society of Regional Anesthesia is creating standards for those facilities run primarily by anesthesiologists.

By and large, the standards address the number of specialists involved in patient care and their experience, education, and degree of participation in the subspecialty of pain management. Other criteria include the breadth of pain problems evaluated in the unit, the variety of therapeutic modalities offered, and the hours the facility functions.

## Essential Components of the Pain Management Team

The staff of the pain clinic must have a genuine interest in dealing with pain patients, an appreciation for how acute pain is fundamentally distinct from chronic pain, and the time for a thorough, systematic evaluation of the patient (2, 3, 5). Most chronic pain patients present a complex management problem. Their pain behaviors and individual distress are frequently out of proportion to the physical findings. The long-standing suffering of these patients often results in dysfunctional behavior patterns and significant life disruption. As a result, pain with its many manifestations and repercussions needs to be addressed by a multidisciplinary team. A purely physiologic approach will ignore a critical part of the picture. The routine of single-specialty management may contribute to the perpetuation of chronic pain suffering because it does not address the psychosocial, occupational, and familial consequences of the pain in the patient's life. Any specialist involved in the management of pain patients must appreciate problems beyond his or her area of specialty training.

There are a number of essential components of the pain management team (1–6), which can be divided into several groups:

- Those directly treating the patient include physicians, psychological counselors, nurses, patient managers, and kinesiologists.
- Consultant services, such as occupational therapy, rehabilitation specialists, physical therapy, dietary specialists, social services, and vocational rehabilitation, should also be readily accessible.
- Office support personnel, such as a business manager, an administrative coordinator, a receptionist, and secretarial staff, are indispensable contributors to a smoothly functioning team.

### Direct Treatment Personnel

**Physicians.** A physician is likely to be the director of the pain clinic. He or she may have both clinical and administrative responsibilities, as may assistant directors. Physicians take a history and perform a physical examination in an attempt to define the organic parameters of the pain as well as ascertain the extent of the physical limitations imposed by it. They may also recommend specific tests that complete the diagnostic data base. They prescribe appropriate medications, perform nerve blocks or invasive therapies, as well as coordinate other aspects of the patient's total care.

An additional vital function is that of educator. Patients must be taught about the etiology of their pain through explanation, pictures, and models. It is the physician's job to determine if the primary cause of pain is benign or malignant; to highlight the differences between acute and chronic pain; and to clarify the rationale for each modality that is recommended in the treatment program. Furthermore, the physicians must communicate effectively with referring physicians so that there is continuity of pain care in the home as well as an educational benefit to the community physician.

**Psychological Counselors.** Psychological services may be provided by clinical psychologists, psychiatrists, or pain counselors (1, 3, 4, 6). All need to have a working knowledge of the psychodynamics of chronic pain and must be well versed in dealing with aberrant coping styles and dysfunctional behavior.

Many patients who have chronic pain are not psychologically sophisticated and tend to be concrete in their thinking. As a result, they express psychological distress through pain behaviors and are less likely to respond to long-term, insight-oriented psychodynamic therapy. Pain counseling needs to focus on improving self-esteem, encouraging behavior modification, and augmenting coping mechanisms because all "pain" may not be completely removed with medical therapy.

These counselors also need to identify the rare patient who is severely dysfunctional, flagrantly psychotic, suicidal, or homicidal, and who is then more appropriately managed in an inpatient psychiatric setting.

Finally, psychological specialists can help identify those patients who are so rigidly attached to their pain by primary or secondary gain that they are unlikely to respond to any treatment until this critical link is adequately addressed. It is important to identify these patients because the risk-to-benefit ratio of invasive procedures is often unacceptable. Because these patients "need" or "want" their pain, they may not admit therapeutic benefit from nerve blocks or surgery. Unfortunately, they are not mute about any complications, real or imagined. In fact, a negative outcome often gives further support to their view of themselves as disabled and also fuels their anger with and distrust of the medical system. Realistic expectations about the patient's compliance with therapy are also related to issues of primary and secondary gain.

Both psychologists and psychiatrists are involved in pain clinic work. That one or the other is necessary is very clear because most physicians are not taught about or sufficiently concerned with maladaptive coping styles and behavioral reactions to stress (7). The decision on which specialty is most appropriate for any one patient depends on the needs and demands for psychosocial and psychoeducational services as well as the interests, strengths, and abilities of personnel available to provide pain counseling (1, 3, 4, 6).

Psychiatrists have the most complete understanding of the clinical use of antidepressant and sedative-hypnotic drugs; the use of, or withdrawal from, such drugs may be part of the patient's management. Furthermore, many patients are referred to the pain clinic looking for a more directed approach than that associated with a general mental health center.

Psychologists are beneficial because most patients referred to pain clinics are not psychotic but are suffering from a behavior disability syndrome that is exacerbated by their chronic pain problem. The cognitive and attitudinal aspects of the patient's pain are significant, especially when so much emphasis is placed on the patient's participation in behavior modification and physiological self-regulation techniques. Psychologists often have a broader repertoire of nonpharmacologic approaches to pain management, including self-hypnosis, biofeedback, or relaxation training.

Patients may view a referral to either specialist as suggesting that the pain is "all in their head" or that they are "crazy." However, taking the physical pain away does not immediately solve all the other consequences of pain—the psychosocial, familial, legal, occupational, vocational, or rehabilitation issues. Thus, a careful explanation of the purpose of the psychological consultation is mandatory.

**Nurses.** Nursing personnel facilitate the smooth functioning of the medical evaluation and treatment area of the clinic. They help prepare patients for evaluation, provide assistance with patient movement and disrobing, and check the patient's vital signs and weight. When a nerve block or other diagnostic or therapeutic procedure is indicated, they help position the patient, prepare the equipment, assist the physician, and monitor the patient's recovery.

**Exercise Kinesiologists.** The pain center also needs a person trained in exercise kinesiology who may be a physical therapist, an exercise physiologist, or an athletic trainer. Although more traditional physical therapy with a focus on specific joints and muscle groups may also be necessary, this particular practitioner in the pain clinic milieu must have a holistic approach to patient care and goals that aim toward overall neuromuscular reconditioning and lifestyle changes for the patient that promote fitness.

Once diagnostic studies have failed to show that ongoing pain is related to continued tissue damage, a significant and essential component of treatment and rehabilitation is increased activity, improved fitness, and enhanced functioning. In many patients this requires a gradual but drastic change in their perception of their physical capabilities. Through individual assessment followed by guided instruction in a program of stretching, warmup, workout, and cooldown, the patient can begin to sense progress with physical rehabilitation. This instruction must emphasize the patient's cooperation through the primarily home-based exercise program that is advocated. The exercise specialist will screen patients and readily select those who will best benefit from referral to other fitness-oriented personnel.

### Office Support Personnel

**Administrative Coordinator.** The administrative coordinator meets with each new patient, orients her or him to the facility, and coordinates the day's schedule. In addition, the coordinator is responsible for follow-up phone calls to selected patients. In a teaching hospital, where the resident and attending physicians rotate assignments frequently, it is helpful for the patients to have a consistent, familiar contact in the pain center. This person also supervises the secretarial and receptionist functions and aids patients with billing problems. The coordinator oversees public relations, routine patient processing, and management of the day-to-day facility, housekeeping, and secretarial problems that arise.

**Referral Secretary.** In larger pain clinics, one secretary must necessarily be solely responsible for receiving and processing referrals in an efficient and timely manner, as well as typing reports and correspondence. Businesslike contact with referring physicians is vital to the image and livelihood of the center. The receptionist and schedule manager must have excellent organizational skills to create a schedule that accommodates the requests of patients and specialists alike, all within the time constraints of the work day.

**Business Manager.** The business manager may well be part of the parent department (anesthesiology, neurology, psychiatry) because most pain clinics are not large enough to support this position independently. He or she has the important job of working with insurance carriers, determining acceptable billing codes, and assuring effective billing procedures and collections. Because of the large overhead required by a multidisciplinary pain center and the extended time afforded the evaluation, education, and counseling of these complicated patients, efficient billing and subsequent collection from insurance carriers is basic to the financial survival of the center.

## Facilities and Equipment

The physical arrangement and equipment of the pain center will vary according to the specialists who dominate the pain management team. It is a reasonable goal to have a fixed locus of operation so that the identity of the facility as an established evaluation and treatment unit is secure (1). The presence of a nearly constant staff is desirable because it provides continuity in patient management, recognition, and familiarity between patients and staff (2). The staff members develop rapport, establish cooperative practice patterns, and share a sense of team effort in developing a meaningful treatment plan for each patient.

**Space and Design.** The biggest challenge of the pain clinic is acquiring enough space to allow for the efficient flow of patients in and around the facility. The physical limitations of many patients demand a barrier-free environment.

Inpatient facilities will be best served if they have autonomous space, such as an entire wing of a hospital floor (1).

Outpatient facilities face design challenges that must be addressed (2, 4–6). Because the evaluation and treatment processes are often lengthy, waiting room must be ample and comfort services (bathroom, snack food) must be readily available. The storage of patient records with ready access is necessary, as is space for work stations and related supplies for secretarial and bookkeeping personnel. Private consultation rooms suitable for conferences of small groups (four to six people) are desirable, given the importance of developing patient rapport and confidence through private education of the patient and the family members (who are also affected by the pain). Separate areas for medical, psychological and exercise evaluation and therapy are needed, as is office space for staff members.

Larger facilities will provide conference room space (for up to 20 people) and a library. If inpatients are to be seen in the outpatient pain clinic, hallways and doorways must accommodate stretchers.

A staff lounge area is beneficial to morale and helps to limit recreational conversation in patient care areas that might otherwise disrupt the professional atmosphere of the pain center.

Because no therapeutic modality is without risk, adequate equipment and facilities for the management of complications should prevent any mortality and minimize morbidity in the treatment of pain with invasive techniques. Thus, adequate monitoring devices, oxygen, suction, spot lighting, and interoffice communication equipment are necessary.

## Challenges to the Organizational Structure and Potential Problems

The organization of the pain clinic must allow for efficient collection of records relating to pain problems. The more thoroughly this is achieved initially, the more effective the patient's evaluation and the more comprehensive the treatment planning. The referral process must permit time for patients to complete the pain center's particular questionnaire and screening tools that add to the relevant data base. Ideally, completed records can be previewed before the patient's visit to prompt an outline for approaching problems and determine the need for collateral consultations.

Patient education about the pain center's procedures and goals must begin with an orientation to the facility. Thereafter, the patient is guided sensitively through the day's assessments, and the evaluation concludes with a conference that aims to explain diagnoses and therapeutic options. Patients who understand the therapeutic options and take part in prioritizing their implementation are more likely to participate enthusiastically. Such patients are more likely to feel like a part of the management program than like a victim of the medical system and are more disposed to comply with therapy. Providing the patient with a written summary of the recommendations at the conclusion of the evaluation will encourage timely follow-up by the patient and serve as a common point of reference should problems or questions on treatment subsequently arise. Follow-up by return visit or phone call is arranged as deemed necessary and appropriate for each patient. Documentation of any phone calls and their content is vital to the smooth management of ongoing therapy.

The entire pain center staff must adhere to a limited number of basic principles (maxims):

**First, never challenge the authenticity of or the distress caused by pain.** In this way, care providers and patients start with the same understanding of the pain in an encouraging atmosphere, rather than promoting continued pain behavior and generating an adversarial relationship. Negative laboratory studies do not indicate the absence of "real" pain, nor do they necessarily connote the predominance of psychopathology.

**Second, one must have an open mind about "pain," understanding the significance of personality style, cognitive and affective capacities, personal needs and conflicts, educational level, vocational and social history, and cultural expectations.** Care providers must recognize nonphysical aspects of the complaints as they invariably impact on the expression and manifestations of pain. Specialists must not interpret "pain" within the narrow contraints of their own specialty, while ignoring issues such as secondary gain, personal losses, a past history of abuse, and compensation or pending litigation (1, 3–5, 8).

**Finally, patients and, often, their families need to be actively involved in the treatment process.** Many patients harbor illusions that pain represents some life-threatening disease that remains undiagnosed. Education about the etiology and clinical significance of pain is critical. Acute and chronic pain cannot be effectively treated in the same way. A therapeutic plan for management of chronic pain is based on the hope that when a number of modalities, each contributing a percentage decrease in the pain, are used in combination, significant pain relief will result. In managing pain, one must be careful not to promise "cures." The primary goal is to improve the patient's level of functioning while decreasing as much as possible the frequency and intensity of pain. Also, one attempts to break the cycle of perpetual physician referral, expensive testing, and physical deconditioning brought on by recommended or self-prescribed inactivity.

The goal is to relieve the psychological distress and isolation that so commonly accompany long-standing pain problems by providing the patient with insight into all aspects of pain and encouraging development of coping mechanisms (1–7). To this end, therapy recommendations and their rationale are sent to referring physicians in an attempt to encourage their active involvement. The ultimate goal is the patient's functional rehabilitation.

## Summary

Chronic pain is a complex, multifaceted experience that leads not only to disability and dysfunction but also to suffering and unrealized human potential (2, 6). To be effective, a pain management center must be multidisciplinary in concept as well as practice. It must employ practitioners whose assessments are not restricted to specialty interests. They must communicate effectively with other specialists and cooperate to design and implement the best treatment program for each patient. The ultimate goal is rehabilitation of the patient to a more functional life despite residual pain. This may involve breaking the doctor-shopping cycle for those patients seeking cure, detoxification to more appropriate medications, addressing personal and family conflicts through counseling, assisting the patient in resolving socioeconomic problems, and, ultimately, teaching the patient to function more effectively within the bounds of any physical limitations.

## References

1. Aronoff GM. The role of the pain center in the treatment for intractable suffering and disability resulting from

chronic pain. In: Aronoff GM, ed. Evaluation and treatment of chronic pain. Baltimore-Munich: Urban & Schwarzenberg, 1985:503–511.

2. Ghia JN. Problems in pain management and suggested solutions. In: Ghia JN, ed. The multidisciplinary pain center. Boston: Kluwer, 1988:119–158.
3. Rowlingson JC, Toomey TC. Multidisciplinary approaches to the management of chronic pain. In: Ghia JN, ed. The multidisciplinary pain center. Boston: Kluwer, 1988:45–73.
4. Gottlieb HJ. Multidisciplinary pain clinics, centers and programs. In: Tollison CD, Kriegel ML, eds. Interdisciplinary rehabilitation of low back pain. Baltimore: William & Wilkins, 1989:259–276.
5. Lee VC, Rowlingson JC. Chronic pain management. In: Spilker B, ed. Quality of life assessments in clinical trials. New York: Raven Press, 1990:269–293.
6. Ng LKY. New approaches to treatment of chronic pain: a review of multidisciplinary pain clinics and pain centers. NIDA Research Monograph 36. Rockville, MD: Dept. of HHS, 1981.
7. Fields HL. Pain. San Francisco: McGraw-Hill, 1987:335–343.
8. Howell JD, Lurie N, Woolliscroft JO. Worlds apart. JAMA 1987;258:502–503.

# Instructions for Authors

THE MOUNT SINAI JOURNAL OF MEDICINE, established in 1934 as *The Journal of the Mount Sinai Hospital,* is a peer-reviewed general medical journal. It is indexed or abstracted in *Index Medicus, Hospital Literature Index, International Nursing Index, Excerpta Medica, Chemical Abstracts, Biological Abstracts,* and is available on major databases. Original articles, theoretical explorations, basic science articles, reviews, reports of clinical or research experience, articles on any subject affecting the practice of medicine, short notes in any specialty, letters, and articles by medical students are welcome. The *Journal* is a participant in the Agreement on Uniform Requirements for Manuscripts Submitted to Biomedical Journals.

**Communications.** Address correspondence on manuscript submission and review to: Managing Editor, *The Mount Sinai Journal of Medicine,* Box 1094, 19 E. 98th Street, 4D, New York, NY 10029-6574. Telephone: (212) 241-6108.

**Copyright.** The publisher reserves copyright and renewal rights on all published material, and such material may not be published elsewhere without written permission of the publisher. When submitting a manuscript, include the following assignment of copyright in your cover letter: **In consideration of The Mount Sinai Journal of Medicine (MSJM) taking action and reviewing and editing my (our) submission, the author(s) undersigned hereby transfer(s), assign(s), or otherwise convey(s) all copyright ownership to The Mount Sinai Journal of Medicine in the event that such work is published by MSJM.** Each author must sign the statement.

**Manuscript Preparation.** Submit manuscripts in triplicate, including three sets of illustrations. Manuscripts must be typewritten, **double-spaced throughout (including references, legends, and tables),** on one side of 8½ × 11-in. white bond paper, with generous margins. Number pages consecutively. Begin a new page for title page, abstract, first page of text, references, each table, and legends. **Computer-generated typescripts** must be double-spaced and must be easy for editors and typesetters to read. **Computer tapes or disks:** consult managing editor.

**Title Page.** On the title page, on separate lines (double-spaced), include (1) title of article, phrased as concisely as possible; (2) name of each author, including first name and degree; (3) name and address of departments and institutions from which the work originated; (4) acknowledgments of sources of support; (5) name, place, and date of any conferences at which paper has been delivered; (6) the statement **Address reprint requests to . . . ,** giving full name and address, with zip code, of the appropriate author.

**Abstract.** Original articles should include an abstract, which should be a single paragraph not exceeding 150 words. State the purposes; basic approach or methods; main findings; principal conclusions. Emphasize new and important aspects. When an abstract is provided, a summary at the end is not necessary. In literature reviews and other articles which do not begin with an abstract, the manuscript should conclude with a short summary.

**Key Words.** Supply 3 to 10 key words or phrases at the bottom of the title page, for use in indexing the article. Use Medical Subject Headings from *Index Medicus.*

**Writing Style and Abbreviations.** Authors should aim for conciseness and clarity without repetition. Avoid medical jargon, abbreviated phrasing, and obscure or newly coined abbreviations. Define all abbreviations on first use, except those for standard, universally recognized units of measurement and statistical terms. Spell out such terms as white blood cell, hematocrit, superior vena cava.

**Units of Measurement.** Use S1 units, giving range or interval of normal values; more familiar units may be used in addition.

**Editing.** All papers will be edited to enhance conciseness and clarity without altering meaning and to insure conformity with journal style. The author is responsible for all statements in the article.

**Illustrations.** Photographs, line drawings, graphs, and charts should increase understanding of the text. Line drawings, graphs, and charts should be professionally prepared. All illustrations should be submitted in triplicate as sharp, high-contrast glossy prints. Photomicrographs must be 5⅛ in. wide. On the back of each print affix a gum label with the number of the figure (numbered consecutively in arabic numerals as cited in text), title of manuscript, name of senior author, and the word "top" with an arrow indicating top edge. Illustrations which are to appear together should be mounted accordingly as plates and should correspond to each other in size. Protect the prints by enclosing them in heavy cardboard; do not use paper clips. Photographs of patients can be published only if a copy of the permission is submitted with the photograph. Eliminate names of patients and hospital numbers from x-rays.

**Legends.** Legends should not duplicate the text. Begin legends on a separate sheet of paper, double spaced. Each legend should be numbered to match the illustration it accompanies. Illustrations mounted as plates to appear together should be accompanied by a single legend identifying the separate elements by position ("upper left," and so on) or by letters. If letters are used in the legend, a corresponding letter must appear on the print itself. Identify all abbreviations used in a figure. Indicate magnification and stain for photomicrographs.

**Tables.** Each table should be typed on a separate page, double spaced. Number tables consecutively as they appear in text, using arabic numerals ("Table 1"). Give each table a brief title (for example, "Effect of DMSO on Rats") and type it on a separate line. If abbreviations or symbols are used in the table, identify them in a footnote.

**References.** When citing references in your text, the first work cited is numbered 1, and succeeding references are numbered in sequence; enclose the citation number in parentheses and place it before any punctuation: **Cytomegalovirus (1), steroids (2), and recreational drugs (3) have all been implicated (4).** The reference list includes only works cited in the text, numbered in the order in which they are cited. Type, double spaced, following the general arrangement and punctuation style in the examples below (journal title abbreviations conform to *Index Medicus* style):

1. **Nadelman RB, Wormser GP. A clinical approach to Lyme disease. Mt Sinai J Med 1990;57:144–156.**

2. **International Committee of Medical Journal Editors. Uniform requirements for manuscripts submitted to biomedical journals. N Engl J Med 1991;324:424–428.**

**When a manuscript is accepted for publication, you will be asked to confirm the accuracy of all references.**

**Proofs.** Proofs are sent to the corresponding author from the printer. Stylistic changes which do not alter meaning should not be made at this stage. Because of increased printing charges, the cost of excessive author's alterations beyond routine correction of typographical errors and major errors in data may be billed to the author. Proofs should be corrected and returned to the editorial office within 48 hours.

**Reprint Orders.** A reprint order form is sent to the corresponding author with proofs; return it to the editorial office.

# THE
# MOUNT SINAI JOURNAL
# OF MEDICINE

## Forthcoming:
## Grand Rounds

**Bone Marrow Failure Disorders**
Blanche Alter, MD
Mount Sinai School of Medicine

**Is There A Future for Physicians in America?**
Uwe E. Reinhardt, MD
Princeton University

**Interfaces between Seizures and Affective Disorders: Evolution and Longitudinal Course**
Robert M. Post, MD, Robert L. Finding, MD, and Rene S. Kahn, MD
National Institute of Mental Health

**Interventive Gerontology: Short-Term and Long-Term Possibilities**
Gregory M. Fahy, PhD
Transplantation Laboratory
American Red Cross

**Hypertension and Left Ventricular Function**
Robert Phillips, MD
Mount Sinai School of Medicine

**Enteric Infections**
Gerald T. Keusch, MD
Tufts University School of Medicine

## Theme Issues

### SEPTEMBER 1991
**Diabetes Mellitus in the 1990s**
*Editor:* Sheila Roman

### OCTOBER 1991
**Issues in Medical Ethics:**
**Surrogate Decision Making**
**Drug Use: License vs. Prohibition**
*Editors:* Daniel A. Moros and Rosamond Rhodes

### NOVEMBER 1991
**Asthma: New Concepts**
*Editor:* Alvin S. Teirstein

## Available now:

**Sexually Transmitted Diseases**
Edward J. Bottone, editor
volume 57, no. 4, September 1990
88 pages $15

**Lumbar Disk Disease: Controversies in Neurosurgery**
Leonard I. Malis and Martin H. Savitz, editors
volume 58, no. 2, March 1991
96 pages $15

**The Art and Science of Consultation in Medicine**
Paul L. Gilbert, editor
volume 58, no. 1, January 1991
96 pages $15

*Subscriptions for 1991: $60 individuals, U.S. and Canada, and $65 all libraries; $65 individuals outside U.S. and Canada.*
*To order subscriptions or copies of back issues ($15 each), please send check, payable to The Mount Sinai Journal of Medicine, to the Journal at Box 1094, One Gustave L. Levy Place, New York, NY 10029-6574; counter sales at Suite 4D, 19 E. 98th Street, New York, NY (phone: 212 241-6108; FAX: 212 831-1816).*

3 4805 0348215 1

The Mount Sinai Hospital in New York 1852

Mount Sinai School of Medicine of the City University of New York

# THE MOUNT SINAI JOURNAL OF MEDICINE

VOLUME 58 NUMBER 4 SEPTEMBER 1991

## CONTENTS

### DIABETES MELLITUS: STRATEGIES FOR THE 1990s

**Introduction** *Sheila H. Roman* .......... 273

**Prediction and Prevention Strategies in Type I Diabetes** *George Eisenbarth* .......... 274

**An Approach to the Heterogeneity of Diabetes in Adults** *Sheila H. Roman* .......... 280

**Special Issues in the Management of the Elderly Patient with Diabetes** *Carmen F. Palmer* .......... 287

**Teaching Diabetes Self-Management Skills: A High-Technology Approach** *Patricia Landi Linekin* .......... 293

**Perioperative Glucose Control: Does It Really Matter?** *Alex Stagnaro-Green* .......... 299

**Intensive Insulin Therapy: Euglycemia with a Price Tag** *Sheila H. Roman* .......... 305

### GRAND ROUNDS IN PSYCHIATRY

**Interfaces between Seizures and Affective Disorders: The Uses of Visually Mapping the Evolution and Longitudinal Course of an Illness** *Robert M. Post, Robert L. Findling, and Rene S. Kahn* .......... 310

**Acquired Brain Lesions and Psychiatric Illness: Possible Role of Dopaminergic Systems** *Thomas A. Kramer, Arnold Merriam, Philip Harvey, and Michelle D. Lisoskie* .......... 324

*continued on back cover*

Veterans Administration Medical Center Bronx, New York

BETH ISRAEL MEDICAL CENTER

**For excellent response in the treatment of duodenal ulcers...**

**AXID**® nizatidine

**has the right answers**

- Rapid epigastric pain relief[1,2*]
- Fast and effective ulcer healing[2,3,4]

**AXID**

**PASSES THE ACID TEST**

*Most patients experience pain relief with the first dose.

*See adjacent page for references and brief summary of prescribing information.*

NZ-2943-B-149347 © 1991, ELI LILLY AND COMPANY

**AXID**® (nizatidine capsules)

**Brief Summary. Consult the package insert for complete prescribing information.**
**Indications and Usage:** 1. *Active duodenal ulcer*—for up to 8 weeks of treatment. Most patients heal within 4 weeks.

2. *Maintenance therapy*—for healed duodenal ulcer patients at a reduced dosage of 150 mg h.s. The consequences of therapy with Axid for longer than 1 year are not known

**Contraindications:** Known hypersensitivity to the drug. Because cross sensitivity in this class of compounds has been observed, $H_2$-receptor antagonists, including Axid, should not be administered to patients with a history of hypersensitivity to other $H_2$-receptor antagonists.

**Precautions:** *General*—1. Symptomatic response to nizatidine therapy does not preclude the presence of gastric malignancy.

2 Dosage should be reduced in patients with moderate to severe renal insufficiency

3 In patients with normal renal function and uncomplicated hepatic dysfunction, the disposition of nizatidine is similar to that in normal subjects.

*Laboratory Tests*—False-positive tests for urobilinogen with Multistix® may occur during therapy

*Drug Interactions*—No interactions have been observed with theophylline, chlordiazepoxide, lorazepam, lidocaine, phenytoin, and warfarin. Axid does not inhibit the cytochrome P-450 enzyme system; therefore, drug interactions mediated by inhibition of hepatic metabolism are not expected to occur. In patients given very high doses (3,900 mg) of aspirin daily, increased serum salicylate levels were seen when nizatidine, 150 mg b.i.d., was administered concurrently

*Carcinogenesis, Mutagenesis, Impairment of Fertility*—A 2-year oral carcinogenicity study in rats with doses as high as 500 mg/kg/day (about 80 times the recommended daily therapeutic dose) showed no evidence of a carcinogenic effect. There was a dose-related increase in the density of enterochromaffin-like (ECL) cells in the gastric oxyntic mucosa. In a 2-year study in mice, there was no evidence of a carcinogenic effect in male mice, although hyperplastic nodules of the liver were increased in the high-dose males as compared with placebo. Female mice given the high dose of Axid (2,000 mg/kg/day, about 330 times the human dose) showed marginally statistically significant increases in hepatic carcinoma and hepatic nodular hyperplasia with no numerical increase seen in any of the other dose groups. The rate of hepatic carcinoma in the high-dose animals was within the historical control limits seen for the strain of mice used. The female mice were given a dose larger than the maximum tolerated dose, as indicated by excessive (30%) weight decrement as compared with concurrent controls and evidence of mild liver injury (transaminase elevations). The occurrence of a marginal finding at high dose only in animals given an excessive and somewhat hepatotoxic dose, with no evidence of a carcinogenic effect in rats, male mice, and female mice (given up to 360 mg/kg/day, about 60 times the human dose), and a negative mutagenicity battery are not considered evidence of a carcinogenic potential for Axid.

Axid was not mutagenic in a battery of tests performed to evaluate its potential genetic toxicity, including bacterial mutation tests, unscheduled DNA synthesis, sister chromatid exchange, mouse lymphoma assay, chromosome aberration tests, and a micronucleus test.

In a 2-generation, perinatal and postnatal fertility study in rats, doses of nizatidine up to 650 mg/kg/day produced no adverse effects on the reproductive performance of parental animals or their progeny.

*Pregnancy—Teratogenic Effects—Pregnancy Category C*—Oral reproduction studies in rats at doses up to 300 times the human dose and in Dutch Belted rabbits at doses up to 55 times the human dose revealed no evidence of impaired fertility or teratogenic effect; but, at a dose equivalent to 300 times the human dose, treated rabbits had abortions, decreased number of live fetuses, and depressed fetal weights. On intravenous administration to pregnant New Zealand White rabbits, nizatidine at 20 mg/kg produced cardiac enlargement, coarctation of the aortic arch, and cutaneous edema in 1 fetus, and at 50 mg/kg, it produced ventricular anomaly, distended abdomen, spina bifida, hydrocephaly, and enlarged heart in 1 fetus. There are, however, no adequate and well-controlled studies in pregnant women. It is also not known whether nizatidine can cause fetal harm when administered to a pregnant woman or can affect reproduction capacity. Nizatidine should be used during pregnancy only if the potential benefit justifies the potential risk to the fetus.

*Nursing Mothers*—Studies in lactating women have shown that 0.1% of an oral dose is secreted in human milk in proportion to plasma concentrations. Because of growth depression in pups reared by treated lactating rats, a decision should be made whether to discontinue nursing or the drug, taking into account the importance of the drug to the mother

*Pediatric Use*—Safety and effectiveness in children have not been established

*Use in Elderly Patients*—Healing rates in elderly patients were similar to those in younger age groups as were the rates of adverse events and laboratory test abnormalities. Age alone may not be an important factor in the disposition of nizatidine. Elderly patients may have reduced renal function.

**Adverse Reactions:** Clinical trials of varying durations included almost 5,000 patients. Among the more common adverse events in domestic placebo-controlled trials of over 1,900 nizatidine patients and over 1,300 on placebo, sweating (1% vs 0.2%), urticaria (0.5% vs <0.01%), and somnolence (2.4% vs 1.3%) were significantly more common with nizatidine. It was not possible to determine whether a variety of less common events were due to the drug.

*Hepatic*—Hepatocellular injury (elevated liver enzyme tests or alkaline phosphatase) possibly or probably related to nizatidine occurred in some patients. In some cases, there was marked elevation (>500 IU/L) in SGOT or SGPT and, in a single instance, SGPT was >2,000 IU/L. The incidence of elevated liver enzymes overall and elevations of up to 3 times the upper limit of normal, however, did not significantly differ from that in placebo patients. All abnormalities were reversible after discontinuation of Axid. Since market introduction, hepatitis and jaundice have been reported. Rare cases of cholestatic or mixed hepatocellular and cholestatic injury with jaundice have been reported with reversal of the abnormalities after discontinuation of Axid

*Cardiovascular*—In clinical pharmacology studies, short episodes of asymptomatic ventricular tachycardia occurred in 2 individuals administered Axid and in 3 untreated subjects.

*CNS*—Rare cases of reversible mental confusion have been reported

*Endocrine*—Clinical pharmacology studies and controlled clinical trials showed no evidence of antiandrogenic activity due to nizatidine. Impotence and decreased libido were reported with equal frequency by patients on nizatidine and those on placebo. Gynecomastia has been reported rarely.

*Hematologic*—Fatal thrombocytopenia was reported in a patient treated with nizatidine and another $H_2$-receptor antagonist. This patient had previously experienced thrombocytopenia while taking other drugs. Rare cases of thrombocytopenic purpura have been reported

*Integumental*—Sweating and urticaria were reported significantly more frequently in nizatidine- than in placebo-treated patients. Rash and exfoliative dermatitis were also reported.

*Hypersensitivity*—As with other $H_2$-receptor antagonists, rare cases of anaphylaxis following nizatidine administration have been reported. Rare episodes of hypersensitivity reactions (eg, bronchospasm, laryngeal edema, rash, and eosinophilia) have been reported.

*Other*—Hyperuricemia unassociated with gout or nephrolithiasis was reported. Eosinophilia, fever, and nausea related to nizatidine have been reported.

**Overdosage:** Overdoses of Axid have been reported rarely. If overdosage occurs, activated charcoal, emesis, or lavage should be considered along with clinical monitoring and supportive therapy. Renal dialysis does not substantially increase clearance of nizatidine due to its large volume of distribution.

PV 2091 AMP
[091190]

**References**
1. Data on file, Lilly Research Laboratories.
2. *Scand J Gastroenterol* 1987,22(suppl 136):61-70.
3. *Scand J Gastroenterol* 1987,22(suppl 136):47-55.
4. *Am J Gastroenterol* 1989,84:769-774

NZ-2943-B-149347

*Additional information available to the profession on request.*

Lilly

**Eli Lilly and Company**
Indianapolis, Indiana
46285

When you don't know all
the medicines your patients are taking,
protecting their health
is a crap shoot.

Counseling patients about all their medicines improves their odds of getting well and staying well.

**EVERYONE WINS WHEN YOU TALK**

**Send me a *free* Medicine Counseling Kit.**

Name ______________________

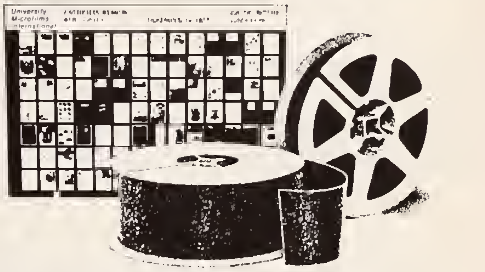
Organization ______________________

Street ______________________

City ______________ State ______ Zip ______

**Mail to:**
National Council on Patient Information
and Education
666 Eleventh Street, NW, Suite 810
Washington, DC 20001
To fax your request -- (202) 638-0773

**this publication is available in microform**



Please send me additional information.
**University Microfilms International**

300 North Zeeb Road
Dept. P.R.
Ann Arbor, MI 48106
U.S.A.

18 Bedford Row
Dept. P.R.
London, WC1R 4EJ
England

Name ______________________
Institution ______________________
Street ______________________
City ______________________
State ______________ Zip ______

# THE MOUNT SINAI JOURNAL OF MEDICINE

The Mount Sinai Journal of Medicine is published by The Mount Sinai Medical Center of New York and has the following affiliates: Beth Israel Medical Center, New York; Bronx Veterans Administration Medical Center, New York; and Elmhurst Hospital Center, New York.

**Editor-in-Chief**

Sherman Kupfer, M.D.

**Editors Emeriti**

David A. Dreiling, M.D.
Lester R. Tuchman, M.D.

**Associate Editors**

Harriet S. Gilbert, M.D.
Julius Wolf, M.D.

**Managing Editor**

Claire Sotnick

**Assistant Editors**

David H. Bechhofer, Ph.D.
Constanin A. Bona, M.D., Ph.D.
Edward J. Bottone, Ph.D.
Jurgen Brosius, Ph.D.
Steven T. Brower, M.D.
Lewis Burrows, M.D.
Joseph S. Eisenman, Ph.D.
Richard A. Frieden, M.D.
Steven Fruchtman, M.D.
Paul L. Gilbert, M.D.
James H. Godbold, Ph.D.
Richard S. Haber, M.D.
Noam Harpaz, M.D.
Dennis P. Healy, Ph.D.
Barry W. Jaffin, M.D.
Andrew S. Kaplan, D.D.S.
Samuel Kenan, M.D.
Suzanne Carter Kramer, M.Sc.
Mark G. Lebwohl, M.D.
Kenneth Lieberman, M.D.
Charles Lockwood, M.D.
Bernard Mehl, D.P.S.
Myron Miller, M.D.
Edward Raab, M.D.
Allan Reed, M.D.
Allan E. Rubenstein, M.D.
David B. Sachar, M.D.
Henry Sacks, M.D.
Robert Safirstein, M.D.
Ira Sanders, M.D.
Martin H. Savitz, M.D.
Clyde B. Schechter, M.D.
Michael Serby, M.D.
Phyllis Shaw, Ph.D.
George Silvay, M.D.
Barry D. Stimmel, M.D.
Nelson Stone, M.D.
Max Sung, M.D.
Rein Tideiksaar, Ph.D.
Richard P. Wedeen, M.D.
Michael F. Wesson, M.D.

**Editorial Board**

Barry Freedman, M.B.A.
Richard Gorlin, M.D.
Nathan Kase, M.D.
Panayotis G. Katsoyannis, Ph.D.
Charles K. McSherry, M.D.
Jack G. Rabinowitz, M.D.
John W. Rowe, M.D.
Alan L. Schiller, M.D.
Alan L. Silver, M.D.
Alvin S. Teirstein, M.D.

*The Mount Sinai Journal of Medicine* (ISSN No. 0027-2507; USPS 284-860) is published 6 times a year in January, March, May, September, October, and November in one indexed volume by the Committee on Medical Education and Publications (Owner), The Mount Sinai Hospital, Box 1094, One Gustave L. Levy Place, New York, NY 10029-6574. Subscription price: individuals, U.S. and Canada, $60 per year; all libraries, $65; individuals outside the U.S. and Canada, $65. Single copies, $15. New subscriptions begin with the first issue of the calendar year. Send notice of change of address 60 days before the change is effective. Second-class postage paid at New York, NY and at additional mailing offices. POSTMASTER: Send address changes to The Mount Sinai Journal of Medicine, Box 1094, One Gustave L. Levy Place, New York, NY 10029-6574. Copyright 1991 The Mount Sinai Hospital.

## *Introduction*

Sheila H. Roman, MD
Chief
Adult Clinical
Diabetes Program
Mount Sinai
Medical Center

# Diabetes Mellitus: Strategies for the 1990s

Diabetes affects approximately one in twenty Americans. It continues to be ranked among the 10 leading causes of death in America and is a significant cause of hospitalization and visits to physicians. This issue of *The Mount Sinai Journal of Medicine* includes adaptations of the talks and workshops from the symposium "Diabetes Mellitus: Strategies for the 1990s," held on April 24th, 1990, and sponsored by the Division of Endocrinology, Department of Medicine, and School of Continuing Education in Nursing at Mount Sinai Medical Center. The symposium provided clinicians with up-to-date theoretical and practical information on caring for the patient with diabetes. I believe physicians will find this compendium of our diabetes symposium useful in their clinical approach to patients with diabetes.

Included in this section are two symposium lectures intended to summarize the state of the art: "Prediction and Prevention Strategies in Type 1 Diabetes" by George Eisenbarth, MD, Head, Section of Immunology at the Joslin Diabetes Center in Boston; and my presentation, "An Approach to the Heterogeneity of Diabetes in the Adult Population."

Also summarized here are the interactive workshops on specific diabetes management techniques given by the staff of our Adult Clinical Diabetes Program: "Special Issues in the Management of the Elderly Patient with Diabetes" by Carmen Palmer, RN, BSN, CDE; "Teaching Diabetes Self-Management Skills: A High-Technology Approach" by Patricia Landi Linekin, RN, BSN, CDE; "Perioperative Insulin Management: Is It Necessary?" by Alex Stagnaro-Green, MD; and my workshop "Intensive Insulin Management: Euglycemia with a Price Tag."

During this decade significant advances will be made in every aspect of diabetes management from diagnosis to treatment. Molecular biology will allow us to precisely understand disease mechanisms and provide the information necessary to develop specific disease markers for the early identification of people at risk for diabetes. Advances in immunology will allow us to safely suppress pancreas-directed autoimmune destructive processes before pancreatic insulin production is disrupted in type 1 diabetes. New therapies will be developed and specifically designed to correct the pathophysiology at the insulin receptor, post receptor, or glucose transporter in type 2 diabetes. The biotechnology explosion will result in the development of glucose sensors, making noninvasive blood glucose monitoring and the development of a "true" artificial pancreas reality. In fact some of these developments are already on the horizon. I look forward with great enthusiasm to being able to incorporate such achievements into the daily clinical care of the patient with diabetes.

Special thanks to Eileen Corkery, RN, MSN, MPH, CDE, for organizing the program and to Alex Stagnaro-Green, MD, associate chief of the Adult Clinical Diabetes Program, for assistance with reviewing and editing the manuscripts.

# Prediction and Prevention Strategies in Type I Diabetes

George Eisenbarth, M.D.

Type I diabetes is an autoimmune disease resulting in pancreatic beta-cell destruction. It is one of the most common chronic childhood diseases. Now, with data gathered worldwide, we can begin to look at a different therapy for type I diabetes—prevention of the pancreatic beta-cell destruction leading to disease. There are two major requirements for prevention of this disease: the ability to accurately identify the prediabetic individual, and the availability of safe and effective preventive therapy.

At the clinical level a number of questions arise. Can one identify individuals prior to onset of diabetes and predict when their diabetes will occur? Although the assays are still somewhat crude at this stage, I believe we can indeed identify individuals at risk and approximate the time of disease onset. Can we prevent diabetes? The trials are just beginning, but it is likely that Type I diabetes can be prevented with some of the immunosuppressive modalities currently available. These are the tasks for this decade.

The data I present from the Joslin Diabetes Center have involved a series of collaborators and excellent postdoctoral fellows, including J. Stuart Soeldner, who began studies of prediabetics when he studied twins at the Joslin Center, and Richard Jackson, who has been instrumental in the development of the dual-parameter mathematical model for prediction of diabetes onset.

Although some differences exist, type I diabetes is an autoimmune disease almost synonymous with insulin-dependent diabetes. For example, there are individuals who have an autoimmune form of pancreatic beta-cell destruction secondary to type I diabetes and who are not yet insulin dependent. Most of these individuals will subsequently develop insulin dependence. There are other rare causes of insulin-dependent diabetes that are not clearly autoimmune, such as poisoning with Vacor (a rat poison that has been taken off the market), and rare genetic syndromes, such as Wolfram's syndrome, in which beta cells are specifically destroyed within the islets.

I use the term "type I diabetes" to refer to an autoimmune, beta-cell-destructive process that can occur at any age and is an HLA-associated disease. It is a disease in which, presumably, T cells are specifically destroying pancreatic beta cells. The predictability of type I diabetes hinges on the fact that this disease is not a disease of acute onset, as was previously believed, but rather develops slowly over the course of several years. The body's pancreatic islet cells are gradually destroyed, until, when approximately 80% of the beta cells are gone, the disease manifests itself clinically. As progressive numbers of beta cells are destroyed, the body is eventually deprived of sufficient insulin production to allow adequate cellular glucose transport. This results in the sequelae of "sudden onset": polyuria, polydipsia, polyphagia, and ketonemia. There are babies who can develop type I diabetes in a year or so, and I think we now have the tools to predict who is going to develop type I diabetes in a year, and who is not going to develop this disease for as much as 15 to 20 years. Development of type I diabetes can be conceptually divided into a series of stages (Fig. 1).

---

Adapted from the author's presentation at the symposium "Diabetes Mellitus: Strategies for the 1990s" held at The Mount Sinai Medical Center on April 24, 1990. From the Section of Immunology, Joslin Diabetes Center, Boston. Address reprint requests to the author, Associate Professor of Medicine and head of the Section of Immunology, Joslin Diabetes Center, One Joslin Place, Boston, MA 02215.

### Genetic Predisposition

In this first stage of genetic susceptibility, autoimmunity has not yet been triggered. There are no autoantibodies to pancreatic antigens. Blood sugar levels, results of the intravenous glucose tolerance test, and insulin secretion are all normal. Why are these individuals considered at risk? Family studies have shown that 5%–10% of the time these individuals have a first-degree relative with type I diabetes. The risk in the general population for developing type I diabetes is about 1 in 300 for the white population. An individual having a diabetic child or a first-degree relative with type I diabetes has about a 1 in 20 chance of developing type I diabetes. If a father has type I diabetes, 6% of his progeny have a chance of developing the disease, and a woman who is diabetic has a 1%–2% chance of having diabetic offspring if she is diabetic during pregnancy. An identical twin with a diabetic twin has a 30% to 50% chance of becoming concordant for type I diabetes. Twins give the limit for the genetic penetrance of the disorder. Nonidentical twins have a 5% risk, similar to that of a nontwin sibling, suggesting that acute environmental effects might not be important in determining which twin develops diabetes. If both parents have type I diabetes, and current series are small, the risk for diabetes in offspring may be as high as that for an identical twin, 30%–50%.

These studies suggest that the number of genes that complement each other to create type I diabetes is relatively small. In animal models, different strains of diabetes-prone mice have different genes creating susceptibility. For instance, if one crosses an NOD mouse that develops type I diabetes with a db mouse that also develops a form of diabetes, none of the offspring will develop diabetes. Therefore, one would presume that if there is a high risk of developing diabetes in the offspring of two type I diabetic individuals, the genes that create diabetes susceptibility in humans are complementary.

These data demonstrate the empiric risk of developing diabetes without the knowledge supplied by laboratory testing. One can do much better in predicting genetic susceptibility with specific tests, since the location of one of the set of genes creating diabetes susceptibility is known and well studied. Essentially everyone who develops type I diabetes has genes within the HLA region of the sixth chromosome that are associated with type I diabetes and, in a family in which one child with type I diabetes inherits the A HLA region from his father and the C HLA region from his mother, a sibling who is HLA identical to the diabetic child (A and C) will have a 1 in 5 risk of developing diabetes. If the sibling is fortunate and inherits neither of the diabetes-risk HLA regions, then the risk of developing diabetes is about 1 in 100. Since the risk in the general population for developing type I diabetes is about 1 in 300 for the general white population, this demonstrates the tremendous influence of the HLA region on susceptibility to diabetes.

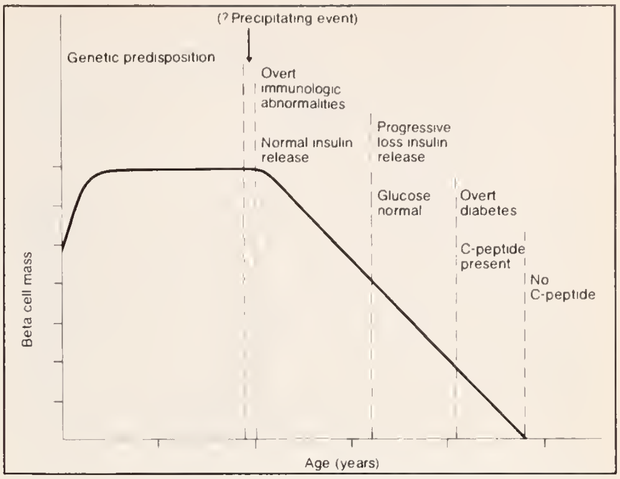


**Fig. 1.** Hypothetical stages in development of type I diabetes. *Reproduced* with permission from Eisenbarth GS, N Engl J Med 1986; 314:1360–1368.

**Genes that Create Susceptibility.** What are the actual genes within the HLA region that create susceptibility? This is a rapidly developing area. The HLA region is approximately 3 million base pairs long and has room for more than 100 genes. The major histocompatibility complex region of the sixth human chromosome is an essential region in terms of genes that affect the immune system, in particular the class II or immune-response genes HLA-DP, HLA-DQ, and HLA-DR. Most current research in diabetes susceptibility focuses on the DQ gene of this HLA region. A caveat: many genes in this region can change susceptibility, and the story is not as simple as DQ beta genes underlying susceptibility for all haplotypes, but polymorphisms of the DQ beta chain probably underlie the susceptibility for two HLA haplotypes: the DR4 haplotype, which is positively associated with diabetes, and the DR2 haplotype, which protects from diabetes.

The DQ molecules are made up of an alpha and a beta chain, both of which are polymorphic and are expressed on the surface of antigen-presenting cells. Variations in the amino acid sequence of the terminal portions of these chains influence the immune response to specific antigens (hence immune-response genes). Ninety-five percent of individuals who develop type I diabetes have HLA-DR3 or HLA-DR4. Most individuals who develop type I diabetes and are HLA-DR4 have a specific sequence of their DQ beta gene (termed DQw8 or DQ beta 3.2 or DQBI*O302) that is associated with diabetes. Individuals who type HLA-DR2 have a particular DQ beta sequence associated with resistance to diabetes (DQBI*0602). Sequences of the DQ beta gene predict susceptibility in DR4- or resistance in DR2-bearing haplotypes. Exceptions help to prove the rule: for these haplotypes, DQ beta genes determine susceptibility. For example, we (1) have studied a family with three HLA-identical siblings who developed type I diabetes, and all type as DR2 and DR1. Since HLA-DR2 usually protects from diabetes, why did these three children develop type I diabetes? Sequencing the DQ beta genes and alpha genes of the members of this family showed that their DQ beta gene sequence was unique and differed from the protective DQ beta sequence usually associated with DR2, whereas their DR beta genes had the usual protection-associated sequence.

For DR2 and DR4 haplotypes, there is a good correspondence with specific molecular sequences of the DQ beta gene and diabetes susceptibility. One of the sequence differences that has correlated with diabetes is lack of aspartic acid at position 57, originally reported by Hugh McDevitt (2). This finding is likely only part of the story. These molecules function as a whole molecule, and many different amino acid changes conferring increased susceptibility are possible.

In general, the function of these immune response genes is beginning to be understood. An antigen-presenting cell, for example a macrophage, will take up an antigenlike insulin and cleave it into small peptides. A given peptide fragment is then bound within a cleft of the macrophage's immune response molecule (that is, DQ molecules) and the antigen presented to T cells. T cells require presentation of an antigen by such histocompatibility molecules because they cannot bind to an antigen that is free in solution. Therefore, these class II molecules, whose sequence variation is helping to predict type I diabetes, function largely to present antigen to T lymphocytes.

We have thus reached the stage at which individuals in the general population can be identified as carrying the sequence DQw8 (DR4-related) on one chromosome and DQw2 (DR3-related) on the other chromosome. If one carries both of these genetic sequences, the risk for developing diabetes, even in the general population with no diabetic relative, is about the same as that of a sibling of a diabetic (approximately 1 in 15). Such individuals make up 40% of those who develop type I diabetes, but only 3% of the general population. Thus one can currently take buccal smears from a large population and identify those individuals with a significant genetic risk of type I diabetes.

What are these genes doing? One hypothesis is that they present specific islet antigens. If they are presenting antigen, what antigen are they presenting? We would like to propose the hypothesis that the antigen being presented for the DR4 haplotype is insulin. We find a strong correlation with an immune response to insulin, namely insulin autoantibodies, and DR4 haplotypes. We find no influence of DR3 haplotypes on insulin autoantibodies, and no influence of DR4 on other autoantibodies.

Thus far, I have discussed the DQ gene for DR4 and DR2 haplotypes. Another haplotype strongly associated with diabetes is DR3. Type I diabetes is often associated with a whole series of autoimmune diseases that are also associated with DR3 (3). The DR3 allele is part of an extended haplotype including B8, DR3, and certain complementary alleles associated with type I diabetes. This haplotype is associated with Addison's disease, Graves' disease, myasthenia gravis, and celiac disease. We occasionally see patients who have all of these diseases, and their first-degree relatives are likely to have other autoimmune diseases. We do not know what gene on the DR3 haplotype creates susceptibility, because DQ sequence variation does not subdivide DR3-expressing haplotypes relative to diabetes risk.

On which cell do these genes act to create susceptibility? We do not know in humans, but in animal models of diabetes data is accumulating that these genes create diabetes susceptibility by acting at the level of the bone marrow. On a genetic level, it appears then that the pancreatic islet cells are passive in the development of type I diabetes and the genetic polymorphisms in the animal models that create susceptibility for diabetes act at the bone marrow, either lymphoid precursor cells or macrophages. Transplantation of NOD mouse bone marrow cells into a normal mouse leads to the development of type I diabetes (4). In a similar manner, bone marrow from a normal animal given to an NOD mouse prevented diabetes (4).

## Triggering Type I Diabetes

Relatively little is known about the precipitating or triggering event for type I diabetes. Identical triplets studied over a period of years at the Joslin Center by Soeldner (5) illustrate the problem: The proband developed type I diabetes, the second triplet developed type I diabetes a number of years later, but the third triplet has not and probably will never develop type I diabetes. This third triplet has not activated the autoimmune process after almost 20 years of followup. What determines disease onset?

**Environmental Hypotheses.** One hypothesis is that different environmental factors distinguish the identical twin or triplet who develops diabetes from the one who does not. Congenital rubella is the only environmental factor that is well documented to influence diabetes susceptibility and increase the probability of type I diabetes; but it also increases the probability for thyroiditis and other autoimmune diseases. My own personal bias, which may of course be proven incorrect, is that we are not going to find such an environmental factor.

**Genetic Hypotheses.** There are a number of alternative potential hypotheses at a genetic level to distinguish which identical twin develops diabetes, including the creation of immunoglobulin genes or T-cell receptors by a random process. I would favor such alternative genetic hypotheses for the triggering of type I diabetes. There now are elegant models in the development of hereditary malignancies and antioncogenes, such as for hereditary retinoblastoma (6). In this tumor one abnormal gene leading to malignancy is inherited as a dominant, but a child does not develop retinoblastoma unless in one retina cell the opposite normal allele mutates and is inactivated. Then in that one cell there are two inactive antioncogenes, leading to the proliferation of that one cell as a cancer cell. Such somatic mutations are an excellent biochemical explanation for penetrance. For retinoblastoma one would predict that identical twins would have a concordance rate of about 80%, because the gene is 90% penetrant. If a child with an inherited retinoblastoma gene mutation does not have a second mutational event prior to about age three, the retinal cells will have stopped dividing and the child will not develop cancer. I hypothesize that we need to find an antiautoimmune gene to distinguish the twin who develops type I diabetes from the twin who does not.

**Immunologic Abnormalities.** It is clear that a whole series of autoantibodies are produced prior to development of type I diabetes (Fig. 2). Identified antigens include insulin, carboxypeptidase-H, the antigen of cytoplasmic islet cell antibodies, and a 64K protein that is glutamic acid decarboxylase. For each of those, autoantibodies are created prior to the onset of type I diabetes. Almost no one develops type I diabetes who does not have one or another of these autoantibodies.

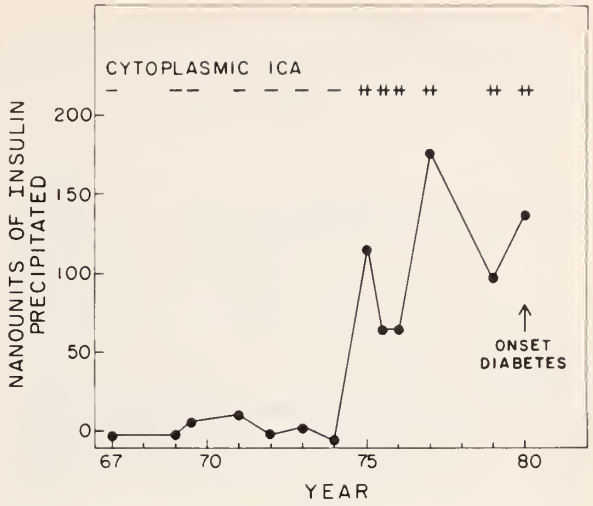


Fig. 2. Activation of autoimmunity in a twin five years prior to development of diabetes. *Reproduced* with permission from Soeldner JS, et al., N Engl J Med 1985; 313:893–894.

The first assay by Doniach and Bottazzo (7) detected cytoplasmic islet cell antibodies (ICA). We think that the target molecule is a complex glycolipid, in the family of a ganglioside, where there is sialic acid and a complex sugar chain. Pancreatic islet cells, in particular, express a monosialoganglioside with mobility between GM1 and GM2 which is either specific or highly selected for the islets and may be the target molecule for ICA.

As a target molecule, insulin is much better characterized. Insulin is comprised of two chains, an A and a B chain. For reasons that are not clear, the insulin autoantibodies of prediabetics do not see random portions of the insulin molecule, and they do not react with the receptor binding domain. The autoantibodies bind to a small determinant of the insulin molecule conserved across most, but not all, species of insulin. I think insulin autoantibodies are the most interesting of the autoantibodies produced. We utilize quantitative radiobinding assays to detect insulin autoantibodies, and they are the only autoantibody thus far whose level correlates with how quickly someone develops clinical type I diabetes (Fig. 3). The level of these autoantibodies correlates with the rate of beta-cell destruction and development of type I diabetes, and it also correlates with the age at which type I diabetes develops. Essentially all children developing type I diabetes by age 5

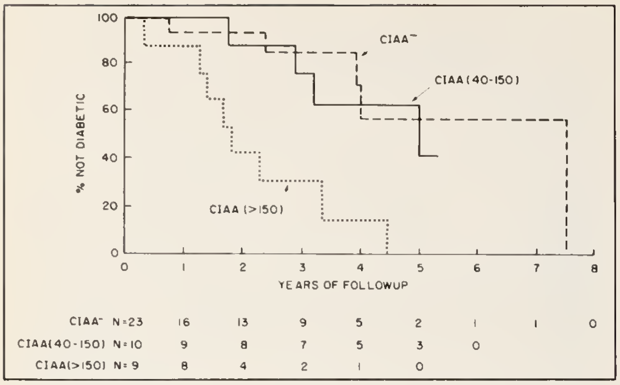


Fig. 3. Influence of level of insulin autoantibodies (CIAA) on progression to type I diabetes in islet cell antibody–positive relatives. *Reproduced* with permission from Ziegler A, et al., Diabetes 1989; 38:1320–1325.

that we have studied have insulin autoantibodies, and the levels at this age are often exceedingly high. The majority of children who develop type I diabetes after the age of 12 to 15 do not produce insulin autoantibodies, and they never have high levels. Once patients receive subcutaneous insulin, they all produce high levels of antiinsulin antibodies, even if they are treated with human insulin.

The number of identified pancreatic antigens is about to expand as the techniques are available to clone unknown antigens. My laboratory has been interested in identifying new target antigens, using expression libraries that are screened with patient sera. One antigen recently identified as an autoantigen is carboxypeptidase-H, an enzyme involved in converting proinsulin to insulin. It is a neuroendocrine molecule, expressed in the adrenal medulla, and it is probably the second most abundant islet granule protein. Insulin is obviously an islet granule protein, and now carboxypeptidase-H, another granule molecule, is found to be an autoantigen. Using our original clone for detection, we find that approximately one-third of pre–type I diabetics make autoantibodies to carboxypeptidase-H.

## Predicting and Preventing Type I Diabetes

**Screening.** For the prediction of diabetes, we currently screen first-degree relatives of type I diabetics for cytoplasmic islet cell antibodies and insulin autoantibodies. Specific autoantibodies (for example, >40 JDF units of ICA, >39 nu/mL of antiinsulin autoantibodies by radiobinding assays) are present in approximately 2% of first-degree relatives.

**Predicting Time of Onset.** The next level in the prediction of diabetes is identification of metabolic abnormalities. Such abnormalities are helpful in predicting the timing of onset of diabetes. Our studies are based on screening approximately 8,000 first-degree relatives for islet cell autoantibodies and approximately 3,000 relatives for insulin autoantibodies. Antibody-positive relatives are followed up to determine who develops diabetes and—more important—when they develop it. The hypothesis we have developed is that the level of insulin autoantibodies correlates with the rate of disease progression, and that first-phase insulin release on intravenous glucose tolerance testing correlates with the degree of beta-cell destruction necessary to produce overt hyperglycemia. Using these two parameters (levels of insulin autoantibodies and first-phase insulin release) in first-degree relatives who are islet cell antibody positive, one can estimate the amount of time until the development of overt diabetes (8).

With these parameters one can obtain broad confidence intervals for the time to onset of diabetes. The data we have indicate that if one is predicted to be diabetic by the model in less than two years, everyone we follow up is diabetic in less than four years. If one looks only at islet cell antibody testing, about 8% per year of ICA-positive relatives with high-titer ICA become diabetic; thus less than half of such relatives are diabetic in four years without the additional secondary criteria of glucose tolerance testing and insulin autoantibodies. Designing an intervention trial based only on ICA would be hard. In addition, some relatives would not develop diabetes for 20 to 30 years or perhaps ever. Almost all ICA-positive relatives identified as having a high risk of developing diabetes using glucose tolerance testing and insulin autoantibodies will become diabetic within three years.

**Potential Preventive Agents and Intervention Trials.** This ability to predict onset of diabetes allows the design of intervention trials in individuals at high risk but before they have overt diabetes. One can then ask whether a particular intervention can prevent type I diabetes with a relatively short, three-year trial.

Many potential agents are available for prevention of diabetes, and in particular, many agents have successfully prevented type I diabetes in either the NOD mouse or the BB rat. A whole series of new drugs is being developed by the pharmaceutical industry. There are now more potential pharmaceutical candidates for preventing diabetes than there are relatives identified as being at high risk for diabetes. To test these pharmaceuticals we need to get on with the process of identifying enough high-risk prediabetic individuals to perform excellent high-quality prevention trials.

Standard immunosuppressive drugs like cyclosporine and azathioprine have been studied in both mouse models and humans (9). These drugs were used after the clinical onset of diabetes and were found to limit beta-cell destruction. They did not, however, result in cure of diabetes for the long term, although C-peptide secretion was preserved while the drugs were administered. Very likely the reason for this failure in long-term cure is that the amount of beta-cell destruction at onset of diabetes is already too severe to permit normal glucose homeostasis.

The first trial in prediabetics that we undertook was with the vitamin nicotinamide because nicotinamide was able to prevent type I diabetes in the NOD mouse model (10). We have, however, abandoned nicotinamide because three individuals we treated with nicotinamide all became diabetic within about a year and a half. A much larger study of nicotinamide is planned in New Zealand.

At the Joslin Center we currently have two approved research protocols. One is for the use of insulin, the other a multicenter trial of azathioprine, in pre–type I diabetics. Any form of immunosuppression has its disadvantages, but I suspect that the disadvantages of azathioprine are fewer than for other immunosuppressive agents. If azathioprine is effective in preventing the progression to diabetes, it would certainly be preferable to becoming diabetic.

Insulin effectively prevents the development of diabetes in NOD mice and BB rats (11); one of the things insulin treatment does is actually remove the ICA antigen from the pancreas. Insulin has a dramatic effect; not only does it prevent diabetes in these animal models, but it prevents insulitis. We have begun trials with insulin therapy in pre–type I diabetics. If insulin therapy was effective, its specificity would make it the therapy of choice. In animal models as much insulin as possible was administered, even to the point of having a subset of animals die from hypoglycemia (9, 11). In humans we are limited to small doses of insulin, on the order of 0.2 units/kg. What we have learned to date is that loss of first-phase insulin release is metabolically regulatable. In the first five children treated with insulin, diabetes has not yet developed after up to two years of followup.

Controlled trials and long-term followup are necessary to evaluate the hypothesis that type I diabetes can be prevented as a chronic autoimmune disease. The prevention of diabetes is an exciting goal of clinical diabetes research. This research is being undertaken without answers to all the basic questions concerning the autoimmunity of diabetes. As basic knowledge advances, more rational therapies for the prevention of diabetes will be developed.

## References

1. Erlich HA, Griffith RL, Bugawan TL, Ziegler R, Alper C, Eisenbarth GS. The identification of type I diabetes siblings with a novel HLA-DQB1 allele and unusual DR2 and DR1 haplotypes implicates specific DQB1 alleles in genetic susceptibility and resistance. Diabetes 1991; 40:478–481.
2. Todd JA, Bell JI, McDevitt HO. A molecular basis for genetic susceptibility to insulin-dependent diabetes mellitus. Trends Genet 1988; 4:129–134.
3. Eisenbarth GS, Jackson RA. The immunoendocrinopathy syndromes. In: Williams textbook of endocrinology (in press).
4. Ikehama S, Ohtsuki H, Good RA, et al. Prevention of type I diabetes in non-obese diabetic mice by allogeneic bone marrow transplantation. Proc Natl Acad Sci USA 1985; 82:7743–7747.
5. Srikanta S, Ganda OP, Jackson RA, et al. Type I diabetes mellitus in monozygotic twins: chronic progressive beta cell dysfunction. Ann Intern Med 1983; 99:320–326.
6. Dryja TP, et al. Homozygosity of chromosome 13 in retinoblastoma. N Engl J Med 1984; 310:550–553.
7. Bottazzo GF, et al. Islet cell antibodies in diabetes mellitus with autoimmune polyendocrine deficiencies. Lancet 1974; 2:1279–1282.
8. Jackson RA, et al. Dual parameter linear model for prediction of onset of type I diabetes in islet cell antibody positive relatives. Clin Res 1988; 36:484A.
9. Bougneres PF, et al. Limited duration of remission of insulin dependency in children with recent overt type I diabetes treated with low dose cyclosporine. Diabetes 1990; 39:1264–1272.
10. Herskowitz RD, Jackson RA, Soeldner JS, Eisenbarth GS. Pilot trial to prevent type I diabetes: progression to overt IDDM despite oral nicotinamide. J Autoimmun 1989; 2:733–737.
11. Godfredsen CF, Buschard K, Frandsen EK. Reduction of diabetes incidence in BB Wistar rats by early prophylactic treatment of diabetes prone animals. Diabetologia 1985; 28:933–935.

# An Approach to the Heterogeneity of Diabetes in Adults

Sheila H. Roman, M.D.

In 1936, Himsworth (1) first suggested that patients with diabetes could be differentiated into insulin-sensitive and insulin-insensitive types. Uncertainty regarding this hypothesis continued until the early 1960s, when Yalow and Berson determined by specific insulin radioimmunoassay (2) that insulin levels, on the average, were higher in patients with nonketotic diabetes than in normals.

**Classification of Diabetes.** In 1979, the National Diabetes Data Group recommended a classification schema (3) to serve as a uniform basis on which to plan and conduct research, collect epidemiologic data, and aid the clinician. This schema separated patients into the diagnostic categories of type 1, or insulin-dependent, diabetes; type 2, or non–insulin dependent, diabetes; and other types of diabetes, for example those associated with pancreatic diseases, hormone disorders, drugs.

Type 1 diabetes is associated with reduced beta-cell mass and absolute insulinopenia; these individuals are ketosis prone and need exogenous insulin for immediate survival. Patients with type 2 diabetes, in contrast, have impaired but often significant insulin secretion, are frequently obese, have insulin resistance at the cell site, and are neither insulin dependent nor ketosis prone.

**Clinical Distinctions and Confounding Clinical Courses.** Although not part of the National Diabetes Data Group classification, certain clinical distinctions between type 1 and type 2 diabetes have been noted repeatedly (Table 1). The type 1 patient is usually characterized by abrupt onset of symptoms, insulinopenia, and dependence on injected insulin to sustain life and prevent ketosis. Type 2 patients are not ketosis prone; more than 80% of them are obese; onset of disease generally occurs after the age of 40.

However, these clinical distinctions are frequently not so clear-cut, as illustrated by the following clinical descriptions of three of my own patients.

**Patient 1,** a 45-year-old woman, is more than 150% above her ideal body weight. After abrupt onset of polydipsia and polyuria, she came to the emergency room in frank diabetic ketoacidosis. She has been treated with insulin.

**Patient 2,** a 32-year-old man with a two-year history of hyperglycemia initially treated with diet and then with oral hypoglycemic agents, is 5% below his ideal body weight. He asked for a second opinion on the need for insulin therapy. At that visit, he was noted to have moderate ketonuria and, on laboratory testing, mild acidosis. He was treated with insulin, but was never precipitously ill.

**Patient 3,** a 50-year-old man of ideal body weight with an eight-year history of type 2 diabetes treated with oral hypoglycemic agents, has never been obese and has achieved good glycemic control with oral agents.

These patients demonstrate the clinical heterogeneity of diabetes in the adult. One of the patients developed classic diabetic ketoacidosis as an obese adult, and the two normal-weight indi-

---

Adapted from the author's presentation at the symposium "Diabetes Mellitus: Strategies for the 1990s" held at the Mount Sinai Medical Center on April 24, 1990.

From the Adult Clinical Diabetes Program, Department of Medicine, Mount Sinai School of Medicine (CUNY), New York, NY. Address reprint requests to Sheila H. Roman, M.D., Chief, Adult Clinical Diabetes Program, Dept. of Medicine, Mount Sinai School of Medicine (CUNY), One Gustave L. Levy Place, New York, NY 10029.

viduals differed markedly in their clinical and therapeutic courses. Clearly, as illustrated by the three patients, diabetes in the adult is not a simple disorder, but rather a complicated metabolic one with marked pathophysiologic heterogeneity. These patients highlight the deficiencies in the present classification schema for diabetes.

Similarly, a recent review of records at a Diabetes Care Center in Connecticut found that 50% of their patients could not be clearly classified (4). This article also questioned the utility of the current classification schema and pointed out the need for a revision. Accomplishing this task for type 2 diabetes can only be achieved through a more thorough understanding than currently exists of the metabolic abnormalities responsible for the disease.

In this article I review the current state of understanding of the pathophysiology of type 2 diabetes, with particular emphasis on the pathophysiologic heterogeneity within this clinical disorder, and I propose a new classification schema.

## Pathophysiology

**Glucose Homeostasis.** Maintenance of normal glucose homeostasis depends on three simultaneous processes that must occur in a coordinated fashion: insulin secretion, stimulation of glucose uptake by the peripheral tissues and the splanchnic tissues, and suppression of hepatic glucose production. After the ingestion of glucose, insulin secretion by the beta cell must effectively promote glucose uptake by the splanchnic and peripheral tissues and also suppress hepatic glucose production in order to maintain normal blood glucose. Defects at the level of the beta cell, muscle, or liver can lead to the development of diabetes. In type 1 diabetes, the defect lies entirely at the level of the beta cell, where autoimmune destruction directly results in insulinopenia. However, in type 2 diabetes, the situation is much more complex. Both impairment of insulin secretion at the beta-cell level and resistance to insulin action at the tissues have been demonstrated.

The obvious consequence of resistance to insulin action at the liver and muscle is hyperglycemia resulting from decreased glucose uptake at the muscle and increased hepatic glucose production. Studies in the Pima Indian population suggest that development of insulin resistance may be the initiating event in type 2 diabetes (5). Pima subjects with a diagnosis of impaired glucose tolerance have been found to have insulin resistance, but normal pancreatic function, whereas Pima subjects with type 2 diabetes have both insulin resistance and pancreatic secretory deficiency. The question, therefore, remains open whether type 2 diabetes can develop without a concomitant disturbance in insulin secretion. Although obese individuals all manifest some degree of insulin resistance and hyperinsulinemia, not all obese individuals will develop diabetes. It is likely that type 2 diabetes develops as a clinical syndrome only after beta-cell function becomes impaired to the point where it can no longer increase insulin secretion sufficiently to overcome the resistance at the liver and peripheral tissues. Subsequently, hyperglycemia develops because of the inability to maintain increased insulin secretion sufficient to overcome the tissue insulin resistance (Fig. 1).

TABLE 1

*Clinical Distinctions between Type 1 and Type 2 Diabetes*

| Clinical feature | Type 1 (insulin dependent) | Type 2 (non–insulin dependent) |
|---|---|---|
| Age at onset | Usually <30 | Usually >40 |
| Weight | Usually lean | 80–90% obese |
| Ketosis | Common | Rare |
| Stability of blood sugar | Extremely labile | Relatively stable |
| Responsiveness to insulin | Sensitive | Variable |

**Insulin Resistance in Type 2 Diabetes and Its Cellular Sites.** What are the potential etiologies of insulin resistance in the pathogenesis of type 2 diabetes, and what are the possible cellular sites of insulin resistance, including those at the insulin receptor, postreceptor, and glucose transporters? A clinical feature commonly observed in patients with extreme states of insulin resistance, including that associated with a variety of genetic syndromes, is acanthosis nigricans. Also observed frequently in type 2 diabetes, this clinical feature in a patient with diabetes suggests insulin resistance and hyperinsulinemia. Insulin resistance and the resulting compensatory hyperinsulinemia potentially initiate a vicious cycle that leads to down-regulation of the number of insulin receptors, exacerbating the insulin resistance.

A variety of potential molecular mechanisms may lead to either quantitative or qualitative defects in the insulin receptor. Evidence for both types of defects has been reported by Taylor et al., using the genetic syndromes of extreme insulin resistance as a model. Utilizing lymphocytes from a patient with one of the genetic syndromes of extreme insulin resistance, the Rabson-Mendenhall syndrome, Taylor (6) was able to show a dra-

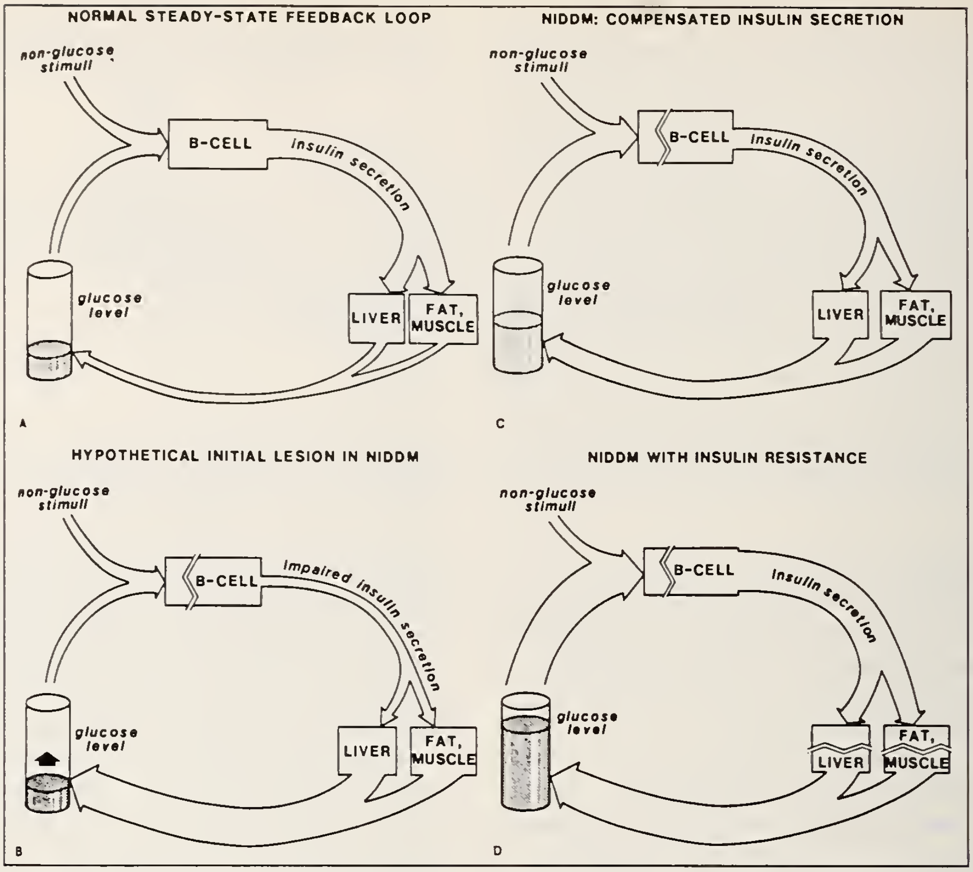

NORMAL STEADY-STATE FEEDBACK LOOP
non-glucose stimuli
B-CELL
insulin secretion
glucose level
LIVER
FAT, MUSCLE
A
NIDDM: COMPENSATED INSULIN SECRETION
non-glucose stimuli
B-CELL
insulin secretion
glucose level
LIVER
FAT, MUSCLE
C
HYPOTHETICAL INITIAL LESION IN NIDDM
non-glucose stimuli
B-CELL
impaired insulin secretion
glucose level
LIVER
FAT, MUSCLE
B
NIDDM WITH INSULIN RESISTANCE
non-glucose stimuli
B-CELL
insulin secretion
glucose level
LIVER
FAT, MUSCLE
D

Fig. 1. (A) Normal basal feedback loop for insulin and glucose. Insulin secretion, through effects on liver, fat, and muscle, modulates blood sugar. Blood sugar level, via interaction with nonglucose stimuli, feeds back to islet to maintain insulin output. (B) Hypothetical initial B-cell lesion of noninsulin-dependent diabetes mellitus (NIDDM). Impairment of insulin output would tend to lead to overproduction and underutilization of glucose, which would cause glucose level to rise. Major reduction of insulin level would only be observed in this situation if glucose level remained normal. (C) Effect of hyperglycemia to compensate for B-cell defect in NIDDM. Insulin deficiency from initial islet lesion leads to hyperglycemia, which in turn stimulates insulin secretion. Thus, resulting basal insulin output is partially restored, and insulin levels are only slightly diminished. As a result of increased insulin output, glucose production and utilization return toward normal. (D) Control of B-cell function in NIDDM when accompanied by tissue insensitivity to insulin, which leads to further hyperglycemia. In response to this further elevation in glucose level, islet may secrete normal or even supranormal amounts of insulin, despite presence of impaired B-cell function. *Reproduced* with permission from Ward WK, Beard JC, Halter JB, Pfeifer MA, Porte D. Pathophysiology of insulin secretion in non-insulin-dependent diabetes mellitus. Diabetes Care 1984; 7:495. Copyright © 1984 by the American Diabetes Association.

matic reduction of insulin binding to cultured lymphocytes from a patient with this syndrome. His data suggested decreased numbers of insulin receptors accounting for the insulin resistance in this syndrome. Autoradiography of solubilized insulin receptors from a normal subject and a subject with the Rabson-Mendenhall syndrome confirmed the finding. Taylor et al. (7) have also shown qualitative defects in insulin receptor binding to changes in pH in cultured lymphocytes from a patient with another genetic syndrome associated with extreme insulin resistance: leprechaunism. The degree to which such intrinsic quantitative or qualitative defects in the insulin receptor are important in type 2 diabetes is uncertain at this time, but these models certainly serve as interesting examples of these defects.

Despite significant advances in the biology of the insulin receptor, the molecular events that couple insulin receptor binding to the regulation of cellular metabolism remain uncertain. Progress in this area has been complicated by the obvious multiplicity of actions of insulin. The well-recognized mechanisms of second-messenger systems, such as cyclic nucleotides, appear not to be central to insulin action. Insulin receptor phosphorylation, with subsequent activation of insulin receptor tyrosine kinase, may represent one important second messenger. Another may be an inositol glycan structure that mimics certain actions of insulin on metabolic enzymes, currently under investigation by Saltiel (8). A variety of specific postreceptor defects in these pathways could account for cellular insulin resistance.

An important result of insulin action, from the point of view of the clinical presentation of the patient, is glucose transport into the cell. A hypothetical translocation model of insulin's action on glucose transport (Fig. 2), taken from Karnieli et al. (9), proposes the following sequence of events: (a) Insulin binds to its specific cell surface receptor, resulting in signal transduction; (b) in response to this signal, glucose transporter proteins from an intracellular pool are translocated to the plasma membrane; (c) following binding and fusion with the plasma membrane, glucose transporters are then exposed to the extracellular medium, giving rise to glucose transport activity; (d) on removal of insulin from its receptor, glucose transporters are translocated back to the intracellular pool. Clearly, defects or alterations in signal transduction, number of intracellular glucose transporters, translocation of these glucose transporters to the cell membrane, or activation of the intracellular glucose transporter pool could result in insulin resistance.

A family of glucose transporters has now been identified and cloned in a variety of tissues. Only the muscle/fat glucose transporter has been shown to be insulin regulatable. Since the muscle/fat transporter is expressed exclusively in tissues that are insulin sensitive with respect to glucose uptake, this protein is an excellent candidate for a highly specific genetic defect predisposing to insulin resistance in type 2 diabetes.

**Deficiency of Insulin Secretion in Type 2 Diabetes.** Potential causes of insulin deficiency in type 2 diabetes include glucose toxicity, intrinsic abnormalities in beta-cell insulin secretion, and islet amyloid polypeptide or amylin. A few reports also suggest that a percentage of patients with clinical type 2 diabetes actually have subclinical and not yet diagnosed insulin-dependent type 1 diabetes. Possibly we can use the dual-parameter model of Eisenbarth to predict when these patients will express their insulin dependency (10).

DeFronzo, Reaven, and other investigators have shown that the relationship between plasma insulin response and type 2 diabetes is complex (11, 12). An individual with a normal fasting glucose would be expected to have a mean plasma insulin concentration of approximately 60 μ/U/mL during the two hours after glucose ingestion. In the setting of impaired glucose tolerance or mild diabetes, the beta cell recognizes that the glucose homeostatic mechanism has become disrupted and can augment its insulin secretory capacity in an attempt to overcome this disturbance. However, once the fasting plasma glucose exceeds approximately 140 mg/dL, the beta cell can no longer maintain its accelerated rate of insulin secretion (11). When the fasting plasma glucose level exceeds about 200 mg/dL, the plasma insulin response to a glucose challenge becomes markedly blunted.

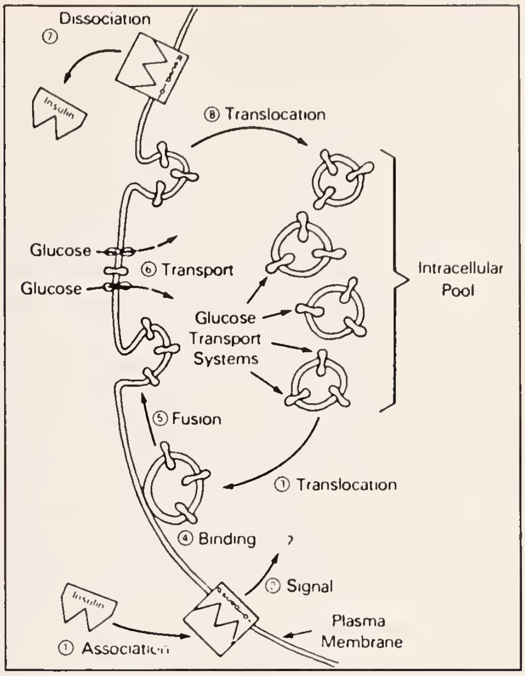


Fig. 2. Schematic representation of a hypothetical mechanism of insulin's stimulatory action on glucose transport in isolated rat epididymal adipose cell. *Reproduced* with permission from Karnieli E, Zarnowski J, Hissin PJ, Simpson IA, Salans LB, Cushman SW. Insulin-stimulated translocation of glucose transport systems in the isolated rat adipose cell. J Biol Chem 1981; 256:4776.

Such observations have led to the concept of glucose toxicity. Consistent with this hypothesis, Kosaka et al. (13) and other investigators have shown improved insulin secretion in obese type 2 patients after improvement in glycemic control. Type 2 patients with poor glycemic control of diabetes can therefore be assumed to have deficient insulin secretion.

Insulin is secreted in response to meals in pulses, two to four pulses occurring after each meal. Polonsky et al. have shown that diabetic individuals have clear abnormalities in insulin secretion. These researchers compared insulin secretory rates for breakfast, lunch, and dinner in control individuals and individuals with type 2 diabetes (14). For controls, sharp peaks of insulin secretion occurred in response to each meal. However, for the diabetic patients, an attenuated insulin secretion response to meals occurred, and peaks of insulin secretion were erratic.

Amyloid deposition has been noted to be increased in the pancreatic islets of patients with type 2 diabetes since the early 1900s. Recent identification by Westermark, Cooper, Clark, and others (15, 16) of a peptide as a major constituent of these pancreatic amyloid deposits has provoked a reexamination of the possible role of pancreatic islet amyloid in the development of type 2 diabetes. Islet amyloid polypeptide, or amylin, appears to originate from beta cells and has been localized to insulin secretory granules in beta cells. Its function is uncertain at this time. Accumulation of amyloid over many years might potentially lead to progressive disruption of the pancreatic islet structure or loss of paracrine communication between cells and possibly result in some of the abnormalities of insulin release seen in type 2 diabetes. Other evidence, however, suggests that amylin's action may be at the skeletal muscle to cause insulin resistance (17).

**Misclassification of Some Insulin-Deficient Type 1 Diabetics as Type 2.** I postulate that a percentage of type 2 diabetic individuals with insulin deficiency are misclassified and are in reality latent adult-onset type 1 diabetics. C peptide is secreted in equimolar amounts as insulin by the beta cell, and fasting C-peptide levels and C-peptide responses to glucagon stimulation have been used as measurements of beta-cell function. Groop et al. compared C-peptide responses to glucagon in ten islet cell autoantibody (ICA)-negative and nine ICA-positive patients at diagnosis, and then again two years later (18). Islet cell antibodies are markers of pancreatic autoimmunity. All patients were clinically considered to have type 2 diabetes at diagnosis and were successfully treated with oral hypoglycemic agents for at least one year. ICA-positive patients showed a progressive and significant decline in their C-peptide response to glucagon over time. The ICA-negative group had good starting levels of C peptide and showed no significant decline. In a follow-up study by Groop et al., examining type 2 patients who were either oral hypoglycemic responders or nonresponders (19), a significantly increased prevalence of ICA and thyroid microsomal autoantibodies was found in individuals who were nonresponders to oral hypoglycemic agents after two years. At the onset, both patient groups had been matched for age at diagnosis of diabetes, to avoid biasing. Furthermore, using life table methods, Irvine et al. calculated an 86% probability of insulin requirement after five years for ICA-positive patients initially treated with oral hypoglycemic agents at diagnosis of their diabetes (20).

**Proposed Classification Schema for Diabetes Mellitus**

Type 1 Diabetes: An autoimmune disease resulting in pancreatic beta-cell destruction.

May manifest:
A. Insulin dependency
B. Insulin deficiency

Type 2 Diabetes: A heterogeneous disorder requiring further subclassification.

May predominantly manifest:
A. Insulin resistance
B. Insulin deficiency

Abnormalities may occur in:
A. Insulin receptor function
B. Postreceptor insulin action
C. Glucose transporters and their metabolic regulation
D. Amylin metabolism or action
E. Pancreatic beta-cell function

My own clinical impression has been that it is not a rare event to see patients who fall into this category. We recently reviewed the cases of diabetic ketoacidosis (DKA) at The Mount Sinai Hospital (21). Of interest, 10% of DKA cases occurred in patients with a previous diagnosis of type 2 diabetes who were being treated with diet or oral hypoglycemic agents prior to their hospital admission for diabetic ketoacidosis. Also, of the patients with new-onset diabetes presenting as DKA, two-thirds were older than 30 years of age. These data suggest that latent type 1 diabetes occurs in a small but significant percentage of individuals initially thought to have type 2 diabetes, and that adult-onset type 1 diabetes is probably not rare. We are currently pursuing the investigation of autoimmune abnormalities in such individuals.

## Diabetes Classification Schema for the 1990s

Ideally, a classification system for diabetes in the 1990s should be based on knowledge of etiology and pathogenesis. It should include a strategy for treatment and subtyping based on current knowledge, which will allow as well for further subdivision to incorporate future findings. My proposed working classification schema is shown in the box.

I consider type 1 diabetes to be an autoimmune disease involving the pancreas. If the autoimmune beta-cell destruction is substantial, insulin dependency results, and in that situation the patient presents acutely, as in the classical, usually juvenile-onset type 1 diabetic. If the autoimmune diathesis at the beta cell is more indolent, the presentation may initially be with insulin deficiency and subclinical type 1 diabetes. Determination of islet cell autoantibodies may serve as a useful specific disease marker for subclinical type 1 diabetes.

In contrast, I view type 2 diabetes as a nonimmune heterogeneous disorder, which will likely require further subdivision as new information on its pathophysiology continues to accumulate. Clinically, either insulin resistance or insulin deficiency may predominate in type 2 diabetes, depending on the underlying pathophysiology. Specific disease markers for the various pathophysiologies are necessary and will undoubtedly be defined in the future.

Lacking such specific disease markers, physicians are confronted with adult patients who when first examined may have had diabetes for many years and also may have defects in both insulin action and insulin secretion. Evidence for predominating insulin resistance in patients includes obesity, acanthosis nigricans, and an elevated fasting C-peptide level. Findings consistent with predominating insulin deficiency include a history of recent significant weight loss, fasting hyperglycemia greater than 200, and the presence of ketonemia or ketonuria. The presence of pancreatic autoimmunity or another autoimmune disease would also be suggestive of insulin deficiency and subclinical type 1 diabetes.

## Treatment Approaches

What, then, do I consider the approach to the adult with diabetes? First, one should assess the degree of hyperglycemia and attempt to determine whether an insulin-resistant or insulin-deficient state predominates. If an insulin-deficient state predominates, I prescribe insulin; if insulin resistance predominates, I prescribe dietary changes, oral hypoglycemic agents, or insulin, depending on the degree of hyperglycemia. When choosing a treatment modality, one should weigh the potential risk of increasing hyperinsulinemia against the benefits of improved glycemia. I always utilize patient education, diet, and exercise as part of the therapeutic regimen. I strive for monotherapy, with either oral hypoglycemic agents or insulin, although in certain clinical situations neither oral hypoglycemic agents nor insulin alone adequately improves glucose control, and combination therapy may be of benefit.

I think it is important to reevaluate therapy at regular intervals, as patients may improve their insulin secretion with improved glycemia and recover from a predominating insulin-deficient state. If glucose toxicity were, indeed, playing a major role, one could make a case for using insulin early for type 2 patients with significant fasting hyperglycemia, without committing the patient to lifelong insulin therapy. Likewise, patients with poor glycemic control on oral hypoglycemic agents may switch from a predominating insulin-resistant state to one of relative insulin deficiency and improve glycemic control substantially with insulin. I believe all patients with positive ICA should receive insulin.

## Summary

Diabetes mellitus in the adult population consists of a group of etiologically heterogeneous disease processes, the common outcome of which is derangement of carbohydrate metabolism. New knowledge in the areas of insulin receptors, insulin action, glucose transport, pancreatic amylin, and pancreatic autoimmunity is likely to lead to the development of disease-process-specific treatment modalities.

### References

1. Himsworth, HP. Diabetes mellitus. Lancet 1936; 1:127–130.
2. Yalow R, Berson S. Immunoassay of endogenous plasma insulin in man. J Clin Invest 1960; 39:1157–1175.
3. National Diabetes Data Group. Classification and diagnosis of diabetes mellitus and other categories of glucose intolerance. Diabetes 1979; 28:1039–1057.
4. Abourizk NN, Dunn JC. Types of diabetes according to National Diabetes Data Group Classification: limited applicability and need to revisit. Diabetes Care 1990; 13:1120–1122.
5. Lillioja S, Mott DM, Howard BV, et al. Impaired glucose tolerance as a disorder of insulin action: longitudinal and cross-sectional studies in Pima Indians. N Engl J Med 1988; 318:1217–1225.
6. Taylor SI. Receptor defects in patients with extreme insulin resistance. Diab Met Rev 1985; 1:171–202.
7. Taylor SI, Roth J, Blizzard RM, Elders MJ. Qualitative abnormalities in insulin binding in a patient with extreme insulin resistance: decreased sensitivity to alterations in temperature and pH. Proc Natl Acad Sci USA 1981; 78:7157–7161.
8. Saltiel AR. Second messengers of insulin action. Diab Care 1990; 13:244–279.
9. Karnieli E, Zarnowski MJ, Hissin PJ, Simpson IA, Salans LB, Cushman SW. Insulin-stimulated translocation of

glucose transport systems in the isolated rat adipose cell. J Biol Chem 1981; 256:4772–4777.

10. Jackson R, Vardi P, Herskowitz R, Soeldner JS, Eisenbarth GS. Dual parameter linear model for prediction of onset of type 1 diabetes, abstracted. Clin Res 1988; 36:585A.
11. DeFronzo RA. The triumvirate: beta-cell, muscle, liver. A collusion responsible for NIDDM. Diabetes 1988; 37:667–687.
12. Reaven GM, Bernstein R, Davis B, Olefsky JM. Nonketotic diabetes mellitus: insulin deficiency or insulin resistance? Am J Med 1976; 60:80–88.
13. Kosaka K, Kuzuya T, Akanuma Y, Hagura R. Increases in insulin response after treatment of overt maturity-onset diabetes is independent of the mode of treatment. Diabetologia 1980; 18:23–28.
14. Polonsky KS, Given BD, Hirsch LJ, et al. Abnormal patterns of insulin secretion in non-insulin-dependent diabetes mellitus. N Engl J Med 1988; 318:1231–1239.
15. Johnson KH, O'Brien TD, Betsholtz C, Westermark P. Islet amyloid, islet-amyloid polypeptide, and diabetes mellitus. N Engl J Med 1989; 321:513–518.
16. Clark A, Lewis CE, Willis AC, et al. Islet amyloid formed from diabetes-associated peptide may be pathogenic in type-2 diabetes. Lancet 1987; 232–234.
17. Leighton B, Cooper GJS. Pancreatic amylin and calcitonin gene-related peptide cause resistance to insulin in skeletal muscle in vitro. Nature 1988; 335:632–635.
18. Groop LC, Bottazzo GF, Doniach D. Islet cell antibodies identify latent type 1 diabetes in patients aged 35–75 years at diagnosis. Diabetes 1986; 35:237–241.
19. Groop LC, Pelkonen R, Koskimies S, Bottazzo GF, Doniach D. Secondary failure to treatment with oral antidiabetic agents in non-insulin-dependent diabetes. Diab Care 1986; 9:129–133.
20. Irvine WJ, Sawers JSA, Feek CM, Prescott RJ, Duncan LJP. The value of islet cell antibody in predicting secondary failure of oral hypoglycaemic agent therapy in diabetes mellitus. J Clin Lab Immunol 1979; 2:23–26.
21. Watts RS, Alvarez-Marfany M, Stagnaro-Green A, Roman SH. Epidemiology of diabetic ketoacidosis in a large metropolitan hospital. Presented at the 14th International Diabetes Federation Congress, June 1991, Washington, DC.

# Special Issues in the Management of the Elderly Patient with Diabetes

CARMEN F. PALMER, R.N., B.S.N., C.D.E.

DIABETES is common among the elderly. Because its prevalence increases with age, there is a direct correlation between the "graying of America" and the number of patients with diabetes. The age-related increase in the prevalence of clinical diabetes and the rapid growth of the elderly population offer challenges to health care providers. Because evidence is increasing that diabetes contributes significantly to mortality and morbidity from both acute and chronic complications in the elderly, the need is heightened for practitioners to more diligently approach the treatment of diabetes in this population.

This review highlights the importance of diabetes care in the elderly population. Some of the factors affecting the management of diabetes in this age group are discussed. The increased potential and contributing factors for hypoglycemia in the elderly are explored. Finally, some recommendations for safe and effective management of diabetes in the elderly are given.

**Demographics.** Approximately 11% of the U.S. population is 65 years of age or older; by the year 2020, an estimated 20% will be in that age group. Ten percent of the population is estimated to have type II diabetes by age 60 years, and prevalence increases steadily with each decade of life. By age 80 years, 16%–20% of people have diabetes (1). Approximately two-thirds of patients hospitalized with diabetes are 65 years or older. In a study of patients 75 years and older, Chalmers and colleagues pointed out that 33% of patients who are hospitalized with diabetes have elevated blood glucose levels which go untreated by physicians 50% of the time (2).

**Importance of Diabetes Care in the Elderly.** Many practitioners assume that diabetes in elderly patients is not a serious problem, because these patients will not live long enough to develop complications. Hence, there tends to be both underdiagnosing and undermanagement of diabetes in this age group. In fact, diabetes in the elderly is not an innocuous, asymptomatic disease. It plays a major role in accelerating morbidity and mortality in patients in their 70s and 80s. The clinical features of diabetes in this age group run the spectrum from apparently benign elevated blood glucose levels to fatal comas induced by hyperosmolar hyperglycemic nonketotic states and diabetes ketoacidosis (DKA). The hypotension, dehydration, acidosis, hyperglycemia, and electrolyte imbalance associated with these severe metabolic states further tax the diminished cardiovascular, pulmonary, renal, and hepatic functional reserves of the elderly patient. Consequently, DKA and hyperosmolar states result in high morbidity and mortality in the elderly. In fact, the mortality from diabetes-related comas rises with age, from 3% in those 50 years or younger to over 40% in patients in their 80s (3).

In addition to these acute complications of diabetes, the elderly are also at high risk for the chronic complications. Since the microvascular complications of diabetes correlate best with the duration of the diabetes and the degree of control of blood sugar, elderly patients are in double jeopardy because they have generally had diabetes longer and their diabetes is generally less well controlled. At the time of diagnosis of type II diabetes, many elderly patients have complications, indicating that the diabetes has gone undiagnosed for a significant period of time.

---

Adapted from the author's presentation at the symposium "Diabetes Mellitus: Strategies for the 1990's" held at The Mount Sinai Medical Center on April 24, 1990. From the Department of Nursing, The Mount Sinai Hospital, New York, NY. Address reprint requests to Carmen F. Palmer, R.N., The Mount Sinai Hospital, Box 1158, One Gustave L. Levy Place, New York, NY 10029.

TABLE

*Pharmacokinetic Properties of Oral Sulfonylureas*

| Drug | Average half-life (hr) | Duration of action (hr) | Dose(s) per day | Route of elimination |
|---|---|---|---|---|
| Tolbutamide | 7 | 6–10 | 2–3 | Hepatic; renal of less active metabolites |
| Acetohexamide | 1.5 | 8–12 | 2 | Hepatic to active metabolite with 5-hr half-life, excreted renally |
| Tolazamide | 7 | 12–18 | 1–2 | Hepatic; renal of less active metabolites |
| Chlorpropamide | 35 | 24–72 | 1 | 80% hepatic; renal of drug and less active metabolites |
| Glyburide | 7–10 | >16 | 1–2 | Hepatic; renal of less active metabolites |
| Glipizide | 2–7 | 12–24 | 1–2 | Hepatic; renal of inactive metabolites |

*Reproduced* with permission from Jackson JE, Bressler R (ref. 6).

**Factors Affecting Management.** Many of the problems encountered in the management of the elderly patient with diabetes are related to the interaction of complications of diabetes (box, left) and normal age-related changes (box, right). A significant consequence of this interaction is the alteration of symptoms of clinical illnesses or of the disease state. One such common alteration of symptoms is the hyperosmolar, hyperglycemic, nonketotic syndrome that elderly persons have on initial examination. This diabetic state is now largely an illness of the elderly that may manifest itself without the polydipsia and polyuria that usually accompany hyperglycemia (4). One-third of elderly patients with this condition were not previously known to have diabetes. For many, their initial clinical symptom is an altered sensorium.

## Hypoglycemia in the Elderly

Hypoglycemia is the most frequent acute metabolic complication encountered in the treatment of diabetes in older patients. A hypoglycemic state that might be well tolerated by a young person can lead to morbidity and even mortality in an elderly person.

A major concern with hypoglycemia in this age group is "hypoglycemia unawareness," or the inability to experience early adrenergic warning symptoms of low blood glucose. This is especially common in the presence of autonomic neuropathy. Hypoglycemia unawareness increases the potential for severe hypoglycemia. In older patients, the combination of vascular compromise, involving such vital organs as the heart and brain, and severe hypoglycemia can be catastrophic.

Major causes of hypoglycemia in the elderly include overmedication with insulin or oral hypoglycemic agents (prescribed or accidental), inadequate food intake, and inappropriate timing of meals with medication action. Significant changes predisposing the elderly to hypoglycemia include impaired renal and liver function, impaired vision, impaired mobility, polypharmacy with drug-drug interaction, depression, cognitive impairment, and social situations.

**Impaired Renal Function.** A major cause of impaired renal function in patients with diabetes is diabetic nephropathy. Its prevalence is 5%–10% after 20 years of diabetes. Renal function also decreases with advancing age, even without underlying kidney pathology. Creatinine clearance decreases nearly 50% between the ages of 20 and 70 years (5). Altered renal function in the elderly patient results in an increased renal threshold for glucose, often 280 mg/dL or more before detectable glycosuria develops. Thus, urine glucose estimation should not be used to determine glycemic control.

Another major consequence of altered renal function is reduced clearance of drugs. This results in prolonged half-life of active drug metabolites and reduced insulin clearance. Thus, increased pharmacologic effects of drugs, including insulin, result in an increased risk of hypoglycemia.

Sulfonylureas must be used cautiously. When these agents are selected for treating elderly patients, their pharmacokinetic properties (Table) must be carefully considered, since these agents and their active metabolites are excreted by the kidneys. Glipizide (Glucotrol) is preferred by many for use with elderly patients (7) because all drug metabolites are inactive and the short half-life of glipizide substantially reduces the risk of prolonged hypoglycemia. Chlorpropamide (Diabinese), on the other hand, is not favored in the elderly because of its metabolism into active products and its prolonged half-life, which increase the risk of prolonged hypoglycemia.

**Impaired Hepatic Function.** Hepatic function decreases with advancing age. Many pharmacologic agents, including all oral sulfonylureas, are metabolized by the liver (Table). Hence, decreased hepatic metabolism is another

**Major Chronic Complications of Diabetes**

Vascular diseases
Macrovascular
- Accelerated coronary atherosclerosis
- Accelerated cerebrovascular atherosclerosis
- Accelerated peripheral vascular disease

Microvascular
- Retinopathy
- Nephropathy

Neuropathic conditions
Sensorimotor neuropathy
- Symmetrical, bilateral
  - Lower extremities (most common)
  - Upper extremities
- Mononeuropathy
- Neuropathic ulcer
- Diabetic amyotrophy
- Neuropathic cachexia

Autonomic neuropathy
- Gastroparesis
- Diabetic diarrhea
- Neurogenic bladder
- Impotence in men
- Impaired cardiovascular reflexes

Mixed vascular and neuropathic diseases
- Leg ulcers
- Foot ulcers

Cutaneous problems

**Major Acute Complications of Diabetes**

Metabolic problems
- Nonketotic hyperosmolar hyperglycemic coma
- Hypoglycemia
- Diabetic ketoacidosis (unlikely in type II diabetes mellitus except during periods of stress such as those caused by trauma and intercurrent infection or disease)
- Lactic acidosis

Infection

*Reproduced* with permission from American Diabetes Association, Clinical Education Program (ref. 9).

**Changes Associated with Aging**

Sensory changes
- Decreased vision
- Decreased smell
- Altered taste perception
- Decreased hearing

Impaired mobility secondary to
- Tremor
- Arthritis
- Other disease processes, e.g., pulmonary, cardiac, cerebrovascular, peripheral vascular

Changes in gastrointestinal function
- Poor dentition
- Decreased salivary secretion
- Delayed gastric emptying
- Altered perception of hunger and thirst

Altered renal and hepatic function
- Altered metabolism of pharmacologic agents

Multiple medical problems

Alteration of symptoms of clinical illness
- Signs and symptoms of hyperglycemia
- Signs and symptoms of infections
- Symptoms of hypoglycemia
- Silent myocardial infarction

Effects of drugs
- Polypharmacy: increased drug interactions and side effects
- Alcohol

Neuropsychiatric problems
- Bereavement
- Depression
- Cognitive impairment and dementia

Social and cultural factors
- Inadequate education
- Long-term dietary habits
- Living alone
- Poverty
- Long-term health beliefs

reason for the cautious use of all oral hypoglycemic agents in the elderly. Insulin degradation, which occurs principally in the liver, is also decreased with increasing age (8).

**Impaired Vision.** Diabetic retinopathy occurs in approximately 50% of all patients with diabetes for more than 10 years. Its prevalence increases with age, from 10% at age 55 to over 30% by age 80 years (9). Diabetic retinopathy is a major cause of impaired vision in the elderly with diabetes. Cataract formation is also a common cause of impaired vision in this age group and cataract formation is dramatically increased in patients with diabetes. Even without retinopathy, impaired vision may be an age-related change. Impaired vision presents the potential for inaccurate insulin dose preparation and a risk of overinsulinization. It also has an impact on the elderly patient's ability to self-monitor blood glucose levels and to prepare food predisposing patients to hypoglycemia.

**Impaired Mobility.** Many conditions can impair mobility in the elderly. Osteoarthritis and cerebrovascular accidents are common causes. Cardiac, pulmonary, and sensory deficits may also contribute to the impairment of mobility in the elderly. The consequences of impaired mobility include inability to prepare and administer insulin correctly, inability to self-monitor blood glucose levels, difficulty with meal preparation,

**Drugs that Enhance Hypoglycemic Activity**

Affect sulfonylurea pharmacokinetics
Displacement from albumin binding site
- Clofibrate
- Halofenate
- Phenylbutazone, oxyphenylbutazone, and sulfinpyrazone
- Salicylates
- Some sulfonamides

Prolongs half-life by interfering with metabolism
- Bihydroxycoumarin
- Chloramphenicol
- Monoamine oxidase inhibitors
- Sulfaphenazole
- Pyrazolone derivatives (e.g., phenylbutazone)

Decreased urinary excretion
- Allopuriol
- Probenecid
- Pyrazolone derivatives (e.g., phenylbutazone)
- Salicylates
- Some sulfonamides

Have their own intrinsic hypoglycemic activity
- Alcohol
- Salicylates
- Guanethidine
- Monoamine oxidase inhibitors
- Beta blockers

**Drugs that Antagonize Sulfonylurea Action and Cause Hyperglycemia**

Affect sulfonylurea pharmacokinetics
Shorten half-life by increasing metabolism
- Chronic alcohol use
- Rifampin

Have own intrinsic hyperglycemic activity
- Acetazolamide
- Beta blockers
- Diazoxide
- Diuretics (e.g., thiazides, furosemide)
- Epinephrine
- Estrogens
- Glucagon
- Glucocorticoids
- Indomethacin
- Isoniazid
- Nicotinic acid
- Phenytoin
- L-Thyroxine

*Reproduced* with permission from American Diabetes Association, Clinical Education Program (ref 9).

and consequent delayed or missed meals, all of which potentiate hypoglycemia.

**Multiple Medical Problems and Polypharmacy.** Elderly patients with diabetes usually have one or more other chronic diseases. At least 50% of these patients have hypertension. Osteoarthritis is frequently present. Heart, sensory, vascular, urinary, foot, and gastrointestinal conditions are twice as common in the population 65 years and older as in individuals 45–64 years old (4). These diseases and their treatments can cause deterioration in blood glucose control, with subsequent hyperglycemia or hypoglycemia (box, left). Thiazide diuretics and nonselective beta blockers used to treat essential hypertension can be particularly problematic. The potassium-wasting effects of thiazides result in impaired release of endogenous insulin and deterioration of blood glucose control. A hazard of the nonselective beta blockers is that they may mask the symptoms of insulin-induced hypoglycemia, thus increasing the risk of severe hypoglycemic reactions.

Because the prevalence of concurrent diseases increases with age, the elderly patient is likely to be taking multiple therapeutic agents. In addition to prescribed medications, the increased use of over-the-counter medications contributes to the problem of polypharmacy. One study showed that elderly women took an average of 5.1 prescription drugs and 3.2 over-the-counter drugs daily.

This excess consumption of medicines is accompanied by an increased risk of adverse medication effect on blood glucose levels. Drug–drug interactions and increased adverse drug side effects may develop as a result of displacement of protein-bound drugs and subsequent enhancement of their pharmacologic and adverse effects. This is a particular problem when treating the diabetes with the first-generation oral sulfonylurea agents. Finally, the potential for error with medication dosages due to complex schedules of multiple drug therapy is always a possibility.

**Depression and Cognitive Impairment.** Twenty percent of elderly individuals suffer significant depression (10). This may be due to loneliness, bereavement from loss of spouse or friends, feelings of uselessness, financial insecurities, and stress related to diabetes or other chronic disease management. Cognitive impairment, ranging from mild age-related impaired intellectual function to severe senile dementia, is present in more than 10% of persons 65 years and older (5).

Both depression and cognitive impairment can alter the ability to take care of oneself. If either is significant, a diabetes regimen requiring significant patient involvement will fail. There will be poor compliance with treatment regimen, inaccurate medication dosaging, poor appetite, and potential for missing meals, with subsequent poor diabetes control and high risk for hypoglycemia.

**Social Factors.** Many social factors may have profound effects on the elderly individual's ability to cope with diabetes management and may influence adherence to the regimen. Poverty, living alone with lack of social supports, faulty notions of health-promoting practice, and lack of appropriate education are all contributing factors. The resultant lack of adherence to the diabetes treatment regimen and lack of consistency in meals and medication will result in significant fluctuations in blood glucose, with increased risk of hypoglycemia and other acute diabetes emergencies.

## General Recommendations for Managing Diabetes Care in the Elderly

In planning care for the elderly, the practitioner must consider the physiologic changes of normal aging, changes due to complications of diabetes, the presence of concomitant illness, the risk of polypharmacy, and psychosociocultural factors. All these factors can have an impact on patients' response to treatment, their increased potential for hypoglycemia and other diabetes-related emergencies, and their ability to learn and safely engage in self-care skills. The following are some general recommendations for the practitioner managing diabetes in elderly individuals:

1. **An interdisciplinary team effort** works best in management of diabetes in the elderly. In addition to the patient and physician, the team includes the patient's family, a diabetes nurse educator, a dietitian, and psychosocial professionals.
2. **A comprehensive multidisciplinary assessment** must evaluate the pathophysiologic changes that may increase the risk of hypoglycemia, patients' ability to learn and their readiness and capability for self-care responsibilities, and the factors that affect self-management and adherence to regimen.
3. **Individualize the treatment regimen** and design it to meet the patient's specific needs.
4. **Establish a goal for optimal blood glucose control.** Blood glucose ranges should be based on the patient's age, general clinical status, and ability for self-management. In establishing blood glucose ranges, the most important aim should be to avoid hypoglycemia. "Safe" rather than "tight" control should be the goal for this age group.
5. **Simplify the regimen** to facilitate compliance and to reduce the risk of errors.
   - Match medication schedules with preexisting schedules as much as possible. This will help the patient to remember medication administration times.
   - Use premixed insulin, for example, Novolin 70/30 or Humulin 70/30, if necessary. These are mixtures of 70% NPH plus 30% regular insulin, which may be used for ease with mixed insulin preparations.
   - Prescribe an insulin dose that is easily measured.
   - Prescribe "low-dose" insulin syringes for ease with insulin preparation, if the patient is taking less than 50 units of insulin per dose. Introduce magnifying devices when necessary.
6. **Self-monitoring of blood glucose levels** is recommended for all insulin-treated patients (11) and for any patient with recurrent hypoglycemia or severe hyperglycemia.
7. **Educate the patient or individual caring for the patient.** Never assume that a patient, because he or she is elderly, does not want to learn or cannot learn. The role of the diabetes educator is paramount in assessing the patient's special educational needs, in educating the patient, and in recommending strategies to overcome certain functional deficits and disabilities.
8. **Always ensure that the patient understands information** being communicated and that the patient can correctly perform survival techniques.
   - Instructions given to patients should always be written. This is important; it is well documented that patients, on leaving the office or clinic, will forget more than 50% of information given verbally.
   - Prior to changing an insulin dose, have the patient demonstrate preparation and administration of the dose being taken. Often the dose verbalized by a patient may not be the actual dose being taken. If a change in dosage is recommended, observe the patient demonstrate preparation of the new dose.
   - Observe the patient's self-monitoring of blood glucose levels and check the interpretation of results.
   - Constantly reassess the patient's skills because motor, visual, and cognitive functioning may deteriorate precipitously in the elderly patient.
9. **Enlist the support of family,** if possible, or of a community agency, if necessary, when a patient demonstrates inability to safely self-manage the treatment regimen.
10. **Frequent follow-up and constant reassessment** of the patient are necessary. Be specific when asking about symptoms of hypoglycemia.
11. **Adjust dosage of hypoglycemic** agents as often as indicated.

## Summary

The elderly patient with diabetes presents unique problems in diabetes management. It is important that practitioners managing elderly patients understand all the factors that impact on diabetes management, in particular those that increase the risk for hypoglycemia and other diabetes emergencies. Given the complexity of the problems encountered, the safe and efficient management of diabetes in the elderly patient requires a team approach.

### References

1. Lipson LG. Diabetes in the elderly: a multifaceted problem. Am J Med 1986; 80(Suppl 5A):3.
2. Chalmers J, et al. Are high blood sugar levels in the elderly ignored? Aust NZ J Med 1987; 17:485.
3. Morley JE. Diabetes in elderly patients. Pract Diabetol 1988; 7(5):6–10.
4. Minaker KL. What diabetologists should know about elderly patients. Diabetes Care 1990; 13(Suppl 2):34–36.
5. Rowe JW. Health care in the elderly. N Engl J Med 1985; 312:827–835.
6. Jackson JE, Bressler R. Clinical pharmacology of sulphonylurea hypoglycemic agents, part 1. Drugs 1981; 22:211–245.
7. Jordan LC. Pharmacokinetic changing in aging. Am J Med 1986; 80(Suppl 5A):31–37.
8. Lipson LG. Diabetes in the elderly: diagnosis, prognosis and treatment. Am J Med 1986; 80(Suppl 5A):10–20.
9. American Diabetes Association. Clinical Education Program. Physicians guide to non-insulin dependent (type II) diabetes, 2nd ed. Alexandria, VA: ADA, 1988:67–83.
10. Weick HH. Senile dementia in geriatrics. In: Platt D, ed. Geriatrics 2. New York: Springer-Verlag, 1982.
11. American Diabetes Association. Consensus Statement: Self-monitoring of blood glucose. Diabetes Care 1991; 14(Suppl 2):57.

# Teaching Diabetes Self-Management Skills:

## A High-Technology Approach

Patricia Landi Linekin, RN, BSN, CDE

DIABETES mellitus is very much a self-care metabolic disorder. Advancing technology has given us an exciting array of tools to assist patients in managing their diabetes. Unless the patient can understand and appreciate the reason for the prescribed treatment, success will not be achieved. The new emphasis on the importance of good blood glucose control for a daily feeling of well being and for the prevention of complications has led to an aggressive treatment approach. This can best be done through a comprehensive, individualized care plan provided by a multidisciplinary health care team.

I present here a historical overview of diabetes self-management and describe current state-of-the-art diabetes technology. The theory of the "teachable moment" and parameters for assessment of a patient's readiness to learn are described. Self-monitoring tools for blood and urine testing as well as devices for insulin administration are explained.

**Historical Perspectives.** Prior to 1921, severe dietary restriction was the only available means of lowering blood sugar. This, of course, was an inadequate therapy, and most patients with type I diabetes died soon after diagnosis. Managing diabetes became a real possibility in 1921 with the discovery of insulin and subsequent commercialization of its extraction procedure from an animal pancreas. These events were quickly followed by the parenteral administration of insulin to people with diabetes. In 1955, oral hypoglycemic medications were developed and proved able to lower blood glucose in people with type II diabetes (1).

Self-assessment of the efficacy of treatment for diabetes began in 1941, when the Clinitest Urine Sugar Analysis Set was introduced. It relied on cupric sulfate reagents, and required the use of a test tube to hold the caustic solution which resulted from the mixture of urine, water, and the Clinitest tablet.

The real breakthrough in self-monitoring came in 1965, when Albert S. Keston (2) developed the technology for the production of TesTape, a simple test strip to detect and measure the concentration of urine glucose. The TesTape contained glucose oxidase which, in the presence of glucose, produced hydrogen peroxide. The second enzyme, peroxidase, catalyzed the reaction of hydrogen peroxide and a chromogen, thereby altering the chromogen and its color in direct relationship to the quantity of glucose originally present. In the mid-seventies, his technology was used in the development of the first test strip for capillary blood glucose reflectance meters. Keston can truly be considered the father of diabetes self-management as we know it today. Although Keston was on the staff of New York University College of Medicine at that time, he came to Mount Sinai Medical Center as a research scientist in 1966.

**Educational Assessment.** An initial assessment of the patient is essential for the diabetes educator to plan an individualized approach to self-management. Ethnic background, education, health-care beliefs, economic status, life style,

---

Adapted from the author's presentation at the symposium "Diabetes Mellitus: Strategies for the 1990's" held at the Mount Sinai Medical Center April 24, 1990. From the Division of Endocrinology and the Department of Nursing, The Mount Sinai Medical Center. Address reprint requests to the author at the Mount Sinai Medical Center, Department of Nursing, Box 1158, One Gustave L. Levy Place, New York, NY 10029.

TABLE 1
*Examples of Application of the Kubler-Ross Model to Diabetes*

| | Patient response to | |
|---|---|---|
| Stage | New diagnosis or long-term management | Onset of complications* |
| Shock and denial | Disregarding consistent food management | Disregarding reduced ability to drive car, particularly in bright sunlight or at night |
| Anger | Refusing to take insulin injection; consciously overeating | Wanting to change doctors |
| Bargaining | Agreeing to test urine, but not recording results | Agreeing to follow regimen only if complications are reversed |
| Depression | Reporting lethargy, fatigue, and feelings of pessimism and helplessness | Not attempting to maximize existing abilities (premature invalidism) |
| Adjustment | Keeping appointments; actively participating in management | Taking advantage of resources, e.g., Bureau of Blind Services |

* Responses described use the example of diabetic retinopathy.

*Reprinted* with permission from Diabetes Mellitus: Diagnosis and Treatment (ref. 3).

eating habits, and support systems must be considered. The use of specific high-technology devices must be weighed with factors such as patient motivation, physical and cognitive abilities, and cost.

A diagnosis of diabetes can be viewed as a loss for a patient. Kubler-Ross's model of stages of acceptance in dealing with a loss can be adapted to describe a patient's initial response to the diagnosis of diabetes (Table 1). Diabetes is regarded by most lay people as a life-threatening, chronic problem. Fear and anxiety can interfere with the learning process for newly diagnosed diabetics who may have witnessed the crippling or fatal effects of diabetes in others, and anticipate these effects in themselves.

The diabetes educator must first assess the level of adaptation of the patient if effective teaching strategies are to be planned. Throughout the period of instruction, the educator should help the patient evolve to the ultimate adaptive level of acceptance, where quality of life issues can be dealt with to achieve a eudamonistic goal of exuberant well-being, and an ability to actualize the self as a person with diabetes.

**Urine versus Blood Glucose Monitoring.** To control diabetes successfully, patients need not only to be educated and motivated, but to have a reliable means of monitoring their own progress so that they can decide when and how to adjust insulin and diet. Urine tests are unsatisfactory for self-monitoring of glucose. They are not only affected by the renal threshold for glucose (which increases with age or compromised renal function); they also do not reflect current blood glucose levels and cannot detect impending low blood glucose. However, if the renal threshold for glucose is adequate, urine glucose testing can be proposed for patients unwilling or unable to perform capillary blood glucose testing.

In contrast, urine ketone testing is an accurate self-monitoring procedure for detection of the initial phase of diabetic ketoacidosis when medical care should be sought.

Self-monitoring of blood glucose has been a major advantage for the person with diabetes. With the advent of blood-glucose meters, people with diabetes were given the opportunity to independently determine an exact blood glucose level at any given time. This provided excellent feedback and became the key to developing a comprehensive, individualized plan for the patient's glycemic control. The resulting awareness of the range of blood glucose fluctuations in a single day is a major incentive for developing increased knowledge regarding diabetes and increased effort in controlling blood glucose. The individual's blood glucose response to stress, illness, exercise, and diet can be evaluated. Hypoglycemic symptoms can be verified and related to a specific blood glucose level, so that appropriate actions can be taken. With proper instruction, some patients can be taught to make daily insulin adjustments based on their blood glucose levels.

Self-testing skills, when correctly performed and understood, can make the person with diabetes a true partner with the diabetes health-care team. As self-testing skills are mastered, patients describe increased feelings of self-confidence and of being "in charge" of their diabetes. The patient should be encouraged to record all tests done at home. The quality of the record-keeping can indicate the patient's interest in glycemic control and provides invaluable information for developing and explaining a therapeutic regimen.

## Blood Glucose Monitoring Technology

Today patients can measure and record their blood glucose several times a day. The American

TABLE 2
*Comparison of Blood Glucose Test Pad Composition*

| Blood glucose meter | Test pad name (type) | Reagent composition |
|---|---|---|
| Reflectance meters | | |
| Accuchek II (5)<br>(Boehringer-Mannheim<br>Indianapolis, IN) | Chemstrip BG<br>(2-zone pad) | Upper zone<br>O-Tolidine 21.5 μg, 3-amino-9 (y-aminopropyl) carbazole-dihydrochloride<br>0.72 μg peroxidase (horseradish)<br>6.0 U glucose oxidase (*Aspergillus niger*) 0.03 U, buffer 85.0 μg, nonreactive ingredients 899 μg,<br>Lower zone<br>tetramethyl benzidine 20.5 μg<br>peroxidase (horseradish) 6.8 U<br>glucose oxidase (*Aspergillus niger*) 0.28 U, buffer 88.0 μg, nonreactive ingredients 757 μg. |
| One Touch (6)<br>(Lifescan<br>Meleptas, CA) | One Touch Blood Glucose Test Strip<br>(1-zone pad) | glucose oxidase, 17 IU; peroxidase, 12 IU; 3-methyl-2-benzothiazolinone hydrazone hydrochloride, 0.06 mg; 3-dimethylaminobenzoic acid, 0.13 mg |
| Enzyme electrode technology | | |
| Exactech (7)<br>(Medisense<br>Cambridge, MA) | ExacTech Blood Glucose Test Strip<br>(1-zone pad) | 3 IU glucose oxidase (*Aspergillis niger*) and 0.1 mg of nonreactive ingredients |

Diabetes Association's 1990 consensus statement addressed some major concerns, noting that individuals with diabetes mellitus as well as health care professionals involved with their care must be trained to understand the purpose of self-monitoring of blood glucose levels, use the technique accurately, and record and make appropriate use of results. The consensus was that the inherent accuracy and precision of self-monitoring devices are generally acceptable when evaluated under carefully controlled situations but added the caveat that in contrast to this general acceptability, in actual patient-care situations the performance of the self-monitoring systems is not always adequate for making clinical decisions (4).

The available technology for capillary blood glucose monitoring consists of test strips, some of which may be used alone, for visual interpretation, as well as for use with a blood glucose meter; and blood glucose meters, which are either reflectance or enzyme electrode meters (Table 2).

**Test Strips.** The composition of the Chemstrip BG (Boehringer-Mannheim, Indianapolis, IN) test pad includes all the chemicals currently used for capillary blood glucose determinations. The plastic test strip contains a reagent pad with two zones.

When blood is placed on the test pads, D-glucose is oxidized to $\gamma$-D gluconolactone, glucose oxidase acting as a catalyst. The hydrogen peroxide resulting from this reaction oxidizes the color indicators in the presence of peroxidase. The intensity of the colors formed is proportional to the glucose concentration in the specimen. The measurement of glucose is determined by visual comparison of both test pads on the strip with color blocks provided on the lot-specific label (5).

Reacted chemstrips can be marked with the date and time and saved in the chemstrip vial for review by the physician and/or diabetes nurse specialist. According to the manufacturer's package insert, "at values of 320 mg/dL or below, colors on strip will be stable for seven days. At values above 320 mg/dL, reacted pads are stable for up to 4 days in the vial" (5).

If a patient or family member is able to correctly interpret visual strips, the strips are an economical option, because chemstrips can be cut in half lengthwise before the application of blood, so that one strip can be used for two tests.

**Meters.** Meter blood glucose determinations are made by processing the information obtained when the blood interacts with the chemicals on the reagent pad.

The first step for both types of meters is similar. When blood is placed on the test pad, the enzyme glucose oxidase on the test strip catalyzes glucose to gluconolactone. The two electrons that are removed from glucose are transferred to a tightly bound flavin adenine dinucleotide on the enzyme. These two electrons are then available for a second reaction. The second step differs.

For reflectance meters, oxygen is reduced to hydrogen peroxidase on the test pad. Hydrogen peroxide is then used to oxidize a dye, changing the dye's color. The color change is measured photometrically or by changes in the amount of specific wavelength of light reflected from the surface. This technology requires optical systems and rigidly timed steps. Recently, meters have been developed with bichromatic instrument optics which read the reacted test strip color without wiping blood from the strips (6).

The enzyme electrode meters measure glucose oxidase activity electrically, using ferrocene, a mediator, to couple the enzyme glucose oxidase to a current-carrying electrode. Electrons produced by oxidation of glucose are attracted to the electrode, and the resulting current is measured. The magnitude of the current is proportional to the amount of glucose in the blood (7).

## Recommendations for Capillary Blood Glucose Monitoring

1. **To be effective,** self-monitoring of blood glucose must be accompanied by a comprehensive education in diabetes self-management.
2. **Visual Chemstrip BGs are adequate** for patients who are unable to afford a blood glucose meter, and who have excellent blue-green color vision. Chemstrips can be dated and stored in the vial for review by physician and/or nurse educator.
3. **Quality must be assured:**
- Control solutions must be used to assure accuracy of meters and strips. Some meters provide test strips to check meter accuracy.
- Meters must be cleaned at least weekly when used daily.
- Meters and strips must be stored appropriately, at temperatures below 86°F (30°C); they should not be refrigerated.

4. **Many factors can affect the results of capillary blood glucose tests:**
- Strips are designed for use with fresh whole blood only. Serum or plasma cannot be used.
- Fluoride preservatives cannot be used for blood sample collection. Fluoride inhibits the enzyme reaction on the test strip. Heparin or EDTA are acceptable as preservatives.
- Extremes in hematocrit (below 30% and above 55%) can cause inaccurate results. For example, a hematocrit of 55% can cause low results in the readings from a One Touch (Lifescan, Inc; Meleptas, CA) or Accuchek II (Boehringer-Mannheim; Indianapolis, IN) meter. High hematocrit affects flow of plasma glucose and oxygen delivery to the test pad (8). This effect may be significant when treating dehydrated patients with hyperglycemic hyperosmolar nonketotic syndrome. Hematocrits less than 30% can cause low results with the One Touch meter and high results with the Accuchek II meter (5). This difference is due to the different design of the test pad: Low hematocrit causes the blood glucose to spread thinly over the One Touch test pad, and causes more plasma glucose to penetrate through to the Chemstrip BG test pad used for Accuchek II.
- Abnormally high levels of ascorbic acid or uric acid may cause low results. Ascorbic acid and uric acid are known reducing agents and therefore interfere with the peroxidase reaction on the test strip. Uric acid above 10 mg/dL (normal 2.6–7.8 mg/dL) and ascorbic acid above 1.2 mg/dL may cause the blood glucose reflectance meters to read lower because the reducing agent competes with peroxide for the chromogenic receptor (9).
- Evaluation conducted with hyperlipidemic patients show that triglyceride levels as high as 6,300 mg/dL have no effect on glucose measurements of Chemstrip BG test strip. However, higher lipemic samples may interfere with this procedure. To be aware of such interferences, baseline glucose values should be established by a clinical laboratory method prior to starting home blood glucose monitoring and rechecked periodically thereafter (8).

5. **Be wary of advertising claims.** *The Wall Street Journal* of March 14, 1990 declared that "only 16% of the nation's 11 million diabetics monitor their blood glucose levels." The race is on to attract consumers. We, as health professionals, must be wary of recommending blood glucose meters which can produce unreliable readings when used in casual settings by our patients.

## Research on Noninvasive Blood Glucose Meters

Active research is in progress employing several technologies for detecting blood glucose levels without requiring a drop of blood.

Biocontrol Technology (Biocontrol Technology, Inc, Indiana, PA) has designed a sensor which "operates by emitting a flow of energy directly into the patient's body tissue. Using advanced software, the sensing device registers numerically how the energy flow interacts with the tissue, indicating the glucose level in the body" (10).

Scientists at the University of Connecticut are developing a "system that would beam light into the eye, where there is fluid containing glucose at about the same levels present in the blood stream" (11). The patient would wear "special glasses equipped with tiny optical fibers aiming an invisible light into an eyeball. The optical fibers could connect to a small sensor located in the glasses or elsewhere." These scientists are basing their study on the way glucose rotates the polarization of light waves. "With the optical fibers, they beam light at blood and then read the resulting shift in the light's polarization" (11).

### Devices for Insulin Administration

Many devices help with insulin administration, and the market is flooded with products that promise much—for example, "pain-free injections," thus appealing to the psychological barrier many patients may have to injecting insulin. Each device must be analyzed for its cost-benefit effect as well as the accuracy of the product claims.

**The Basic Insulin Syringe.** In 1949, Becton Dickinson produced the first disposable insulin syringe. The disposable syringes now come in 1 cc, ½ cc, ⅓ cc, and ¼ cc sizes with 28 gauge or 29 gauge needles. Magnification devices and preset devices are available to assist visually impaired persons in drawing up an accurate dose of insulin. There are products available which can be loaded with a prefilled syringe and set to release the needle into the subcutaneous tissue at the press of a button. An individual assessment of the patient's needs by a diabetes nurse specialist is essential if any of these devices are to be used effectively.

**Insulin Pens.** Insulin pens hold cartridges containing 150 units of U-100 insulin. Cartridges are available for Novolin R, Novolin N, and Novolin 70/30 (Squibb Novo, Princeton, NJ). The dosage can be dialed on the pen. The needle must be manually inserted into the subcutaneous tissue. Site selection is very important with this device, because it is impossible to aspirate for blood prior to injection.

**Jet Injectors.** Jet injectors use pressure to force insulin through the skin surface. They are "needleless." The vial of insulin is placed in the unit, and the dose is dialed. The insulin is injected through the skin with the press of a button. The user must experiment with saline to determine the correct pressure setting for each injection site. A too-shallow injection causes leaking of insulin on the surface of the skin; a higher-than-needed pressure setting can cause extensive bruising. The jet injector must be cleaned and sterilized. Jet injectors are not widely used at this time for insulin, although they are available commercially.

The Amercian Diabetes Association's technical review of jet injectors (12) addresses major concerns with these devices, noting that they:

- are mechanically reliable and accurate, but need assessment by disinterested investigation and published results.
- shorten the total duration of NPH and regular insulin.
- shorten NPH or Lente action and are thus inadequate for dawn phenomenon.
- could denature insulin as a result of its forceful injection through a tiny port, which could lead to an increased incidence of antibody formation.
- have high initial costs; third-party reimbursement is variable.

**Button Infuser.** The Insuflon button infuser device has a button-sized external port connected to a small Teflon catheter that is inserted into the subcutaneous tissue and is taped in place for two or three days. Insulin from a standard syringe is injected through the external port, allowing multiple injections without puncturing the skin each time. As with insulin pumps, clogging of the catheter and infection at the insertion site are possible complications.

**Continuous Subcutaneous Insulin Pumps.** The first insulin infusion pumps were used around 1980. At that time, they were quite large. Today, the pumps are smaller and powered by a rechargable battery. They are worn externally and are often attached to a belt. Insulin is transported from the pump reservoir through a plastic tube to a needle or plastic catheter placed in the subcutaneous tissue. The insertion site is changed every few days. The pump continuously releases regular insulin at a set basal rate. A bolus of insulin is given prior to mealtime by the patient at the press of a button (13). This type of pump is called an "open-loop" pump, referring to the communication loop. The external open-loop pump is the only one available for purchase by an individual. This pump does not make decisions or measure blood glucose. It delivers insulin. The person using it must set the basal rate and decide the appropriate premeal insulin dose based on self-testing of blood glucose. Self-testing must be done four or more times each and every day.

The insulin pumps cost $2,500 to $3,500. The use of the pump requires more intelligence, motivation, judgment, psychological stability, and self-management efforts than more standard methods of insulin delivery. Patients must be carefully selected and educated, because many patients think that pump use is effortless and will eliminate all diabetic complications.

Much research is currently devoted to perfect an implantable insulin pump. An open-loop pump is one version which has been studied. Such pumps are controlled by the patient (as with the continuous subcutaneous insulin pump), but a remote control device is used to determine the rate of insulin delivery for this implantable pump. Also under active investigation is a "closed loop" device that would measure the patient's blood glucose and automatically determine how much insulin to supply—a veritable "artificial pancreas" (14).

The American Diabetes Association position statement on continuous subcutaneous insulin infusion (15) reflects important concerns regarding these devices:

- Patient and professional staff must be willing and prepared to manage this therapy on a 24-hour basis.
- Patients must be carefully selected for strong motivation and ability to use the pump. Self-monitoring of blood glucose levels must be performed, and patients must understand the appropriate response.
- All health professionals should report pump problems to the Centers for Disease Control, the Food and Drug Administration, and the manufacturer for data collection.
- Safety considerations include the fact that undetected interruptions in insulin delivery may result in diabetic ketoacidosis, and the possibility of infections at the needle site.

## Summary

There is, indeed, an explosion of knowledge in the field of diabetes self-management. Special care must be taken in evaluating specific devices before recommending any of them to a patient. In doing so, it is important to consider Owen Edwards' reflections in *Elegant Solutions:* "Aside from nature's creations, there are no perfect solutions. Everything humanity concocts comes with some cost attached. Solutions that seem flawless at first may have repercussions no one ever anticipated" (16).

A multidisciplinary approach is needed to create a diabetes management program to meet the unique needs of each patient. Ongoing education of the patient is essential to assure the best level of diabetes care available.

## References

1. American Diabetes Association. The journey and the dream: a history of the American Diabetes Association. Alexandria, Va: American Diabetes Association, 1990, 2.
2. Keston A. Specific colorimetric enzymatic analytical reagents for glucose (Abstr). In: Abstracts of Papers, 129th Meeting, American Chemical Society, Dallas, Texas, April 1956. Abstr 31c.
3. Davidson MB. Diabetes mellitus; diagnosis and treatment, Vol. 2. New York: John Wiley, 1981, 402.
4. American Diabetes Association. Concensus statement: self-monitoring of blood glucose. Diabetes Care 1990; 13(1):39–40.
5. Boehringer Mannheim Diagnostics. Chemstrip BG: Reagent strips for fast, easy and accurate measurements of blood sugar. 1988 (package insert).
6. Lifescan, Inc. One Touch system technology. In: Reeves-Ellington B, ed. Challenges in diabetes management: clinical protocols for professional practice, New York, NY: Health Education Technologies, Div. of BBDO Health & Medical Communication, Inc, 1988, 73.
7. Medisense, Inc. The ExacTech System. Tec.S., 1989.
8. Gadsen RH. Sources of variation in blood glucose testing. In: Challenges in diabetes management: clinical protocols for professional practice, New York: Health and Medical Publications, 1988, 63–66.
9. Tietz NW, Logan NM. Textbook of clinical chemistry. Philadelphia: WB Saunders, 1986:788.
10. Biocontrol Technology, Inc. Biocontrol technology completes initial noninvasive glucose sensor testing. 1991 news release.
11. Stevens CW. Lab notes: A blood sugar test without drawing blood. Wall Street Journal, April 26, 1990.
12. American Diabetes Association. Technical review: Jet injectors. Diabetes Care 1990; 13(1):39–40.
13. MiniMed Technologies. Modern insulin pump therapy: a guide for successful implementation. Sylmar, CA: MiniMed Technologies, 1987:37–42.
14. Kroll LP, Beaser RS, eds. Joslin diabetes manual, 12th ed. Philadelphia: Lea & Febiger, 1989:172–175.
15. American Diabetes Association. Position Statement: Continuous subcutaneous insulin infusion. Diabetes Care 1990; 13(1):32–33.
16. Edwards O. Elegant solutions. New York: Crown, 1989, 6.

# Perioperative Glucose Control:

## Does It Really Matter?

Alex Stagnaro-Green, M.D.

The benefit of perioperative glucose control is still questioned by many. Proponents argue that hyperglycemia leads to poor wound healing and an increased postoperative infection rate (1, 2). This argument has been countered by many who feel that the risk of hypoglycemia, especially in the hemodynamically unstable patient, far outweighs the benefits of tight glycemic control. Furthermore, no prospective study has been performed that proves that individuals with good glycemic control postoperatively have fewer complications than individuals who remain hyperglycemic.

This article presents evidence for the beneficial effect of "reasonable glycemic control" in the postoperative period. Both animal and human studies are reviewed documenting the deleterious effect of hyperglycemia on both wound healing and the rate of postoperative infections, followed by a discussion of perioperative hormonal changes that result in both insulin deficiency and insulin resistance. This state frequently results in marked hyperglycemia and makes perioperative glucose control difficult. Finally, given the evidence for the beneficial effect of perioperative glucose control, and in light of the hormonal changes that make glycemic control in this setting difficult, I conclude with practical and safe guidelines for perioperative glucose management.

### Effect of Glucose Control on Wound Healing and Infection

**Wound Healing.** For obvious ethical reasons, studies designed to illustrate the effect of glucose control on wound healing need to be performed in an animal model. I have selected two such studies that illustrate the profound effect of hyperglycemia on wound healing. Both studies utilized a mechanical measure of the energy needed to disrupt a healing wound as well as a determination of the collagen (or hydroxyproline) concentration to evaluate the effect of varied glucose control on wound integrity.

Goodson and Hunt in 1977 made rats diabetic by streptozotocin injection and compared the animals' ability to heal following a surgical wound with that of normal nondiabetic rats (3). All the animals had four "stainless-steel wire-mesh cylinders" implanted subcutaneously; after 21 days the surgical wounds were excised. Utilizing a Sanbloom-Muren tensiometer, Goodson and Hunt found that the tensile strength of the wounds of the diabetic animals was 50% less than that of the nondiabetic rats ($p = 0.001$). Furthermore, the diabetic animals had a 50% decrease in the hydroxyproline content of the wound ($p = 0.001$), indicating decreased collagen formation. These authors concluded that "genetically normal animals when made hyperglycemic secondary to insulin deficiency heal less effectively than controls."

Gottrup and Andreassen in 1981 extended the research on wound healing in experimentally induced diabetic animals to its logical conclusion (4). They made 24 rats diabetic with streptozotocin and allowed them to remain hyperglycemic

---

Adapted from the author's presentation at the symposium "Diabetes Mellitus: Strategies for the 1990's" held at the Mount Sinai Medical Center on April 24, 1990.

From the Adult Clinical Diabetes Program, Department of Medicine, Mount Sinai School of Medicine (CUNY), New York, NY. Address reprint requests to Alex Stagnaro-Green, M.D., Assistant Professor and Associate Chief, Adult Clinical Diabetes Program, Department of Medicine, Mount Sinai School of Medicine (CUNY), Box 1055, One Gustave L. Levy Place, New York, NY 10029.

(experimental diabetes). They induced diabetes in another 22 animals with streptozotocin but returned them to euglycemia with frequent insulin injections (insulin-treated diabetes). A third group consisted of nondiabetic control animals (control). All animals underwent abdominal surgery and were killed at either 7 or 20 days postoperatively; then the surgical wound was excised. No differences in breaking strength (maximum force required to disrupt the wound) or breaking energy (energy needed to disrupt the wound) were found between the control group and the insulin-treated diabetes group. On the other hand, the animals who remained hyperglycemic (experimental diabetes) displayed a decrease in both the breaking strength and breaking energy needed for wound disruption. The experimental diabetes group also had decreased collagen concentration and content compared to both the control group and the insulin-treated group. The authors concluded that "diabetes impaired the strength in wounds . . . and . . . normalization of the blood glucose by insulin treatment prevents this effect of diabetes."

**Ability to Fight Infection.** Both in vitro and in vivo evidence for the deleterious effect of hyperglycemia on the ability to fight infection can be found in the literature. Mowat and Baum in 1971 examined the chemotactic ability of polymorphonuclear (PMN) leukocytes in both diabetics and normals (5). They found that the chemotactic index of the leukocytes was markedly reduced in individuals with diabetes ($p < 0.0005$). This defect of chemotaxis was abolished when the cells were incubated with insulin prior to testing. Consequently, their study revealed a reversible abnormality in chemotaxis in the diabetic.

Bagdade et al. in 1972 described another reversible defect in the white blood cells of individuals with diabetes (6). The researchers evaluated the in vitro ability of the PMN leukocytes of the uncontrolled diabetic to phagocytize and kill a standard inoculation of type 25 pneumococci. They found that the leukocytes from the patients with uncontrolled diabetes had both a decrease in the rate of bacterial killing ($p < 0.01$) and a decrease in the ability to phagocytize ($p < 0.001$) compared to controls. These abnormalities resolved after the patients became euglycemic on appropriate therapy. Bagdade et al. concluded that these defects "could contribute to the apparent predisposition of diabetics to bacterial and fungal infection."

Both studies demonstrate marked abnormalities in the infection-fighting abilities of polymorphonuclear leukocytes when the patient is hyperglycemic. Furthermore, a return to euglycemia corrected the abnormalities. These results are, however, based on in vitro experiments, and it can be argued that one should not assume from these findings that glucose control in the postoperative period would decrease the infection rate.

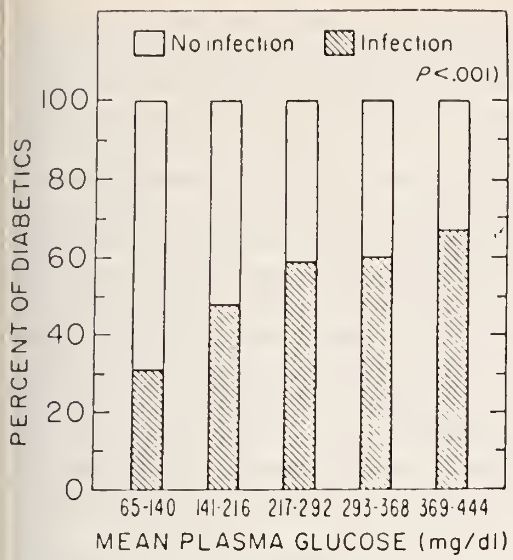
Two studies examining the relationship between hyperglycemia and the rate of infection present evidence in humans for the benefit of glycemic control. Cruse and Foord in 1973 performed a prospective study in which wounds induced at surgical procedures performed at Foothill Hospital in Calgary, Alberta, Canada, were each monitored daily for postoperative infection for 28 days by the same nurse (7). The overall infection rate was 4.75% in the 23,649 surgical wounds that were followed. The infection rate of "clean" wounds (wounds in which the gastrointestinal or respiratory tract was not involved, and where there was no obvious infection or disruption of sterile technique) was 1.8%. However, further analysis of this group revealed that if a patient with a "clean" wound was diabetic, their incidence of postoperative infection increased to 10.7%.

Rayfield et al. in 1982 presented further evidence for the beneficial effect of glucose homeostasis on the rate of infection (8). They conducted a chart review of 241 outpatients. All patients in the study had their serum glucose measured at least three times at a time when they had no obvious infection. By reviewing the patients' emergency room and inpatient records, these authors documented acute infections and correlated them with the mean plasma glucose previously found in the clinic. The results showed a direct correlation between the incidence of infection and mean plasma glucose (Fig. 1). The authors concluded that "the data suggest to us that tight control of blood sugar in the diabetic patient is a desirable goal in the prevention of certain infections."

Thus, evidence is strong that the level of glucose control is directly related to the integrity and strength of a healing wound as well as the rate of postoperative infections. However, to achieve the best glycemic control possible in the perioperative period, an understanding of the factors that make glucose control during this period difficult is needed.

## Hormonal Changes Associated with Surgery

The marked impact of surgery on glucose homeostasis has been known for some time. In 1934



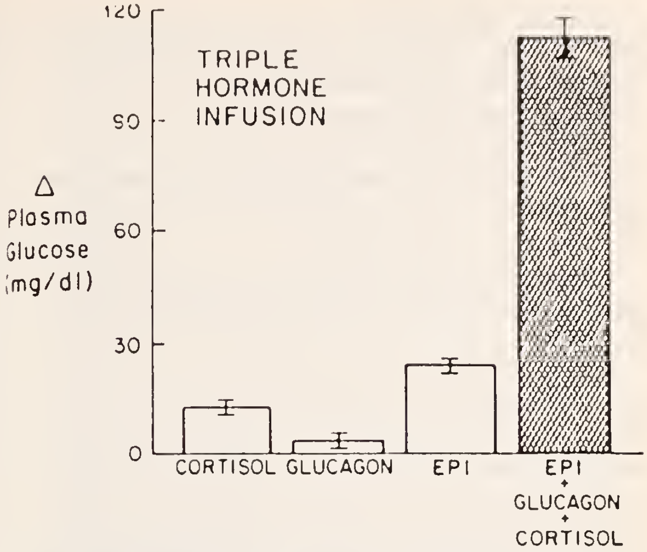


Fig. 1. Left Correlation between overall prevalence of infection and mean plasma glucose levels in 241 diabetic patients ($p < 0.001$). For each patient, three or more fasting glucose determinations were made at times when no evidence of infection existed. *Reproduced* with permission from Rayfield EJ, Ault MJ, Keusch GT, Brothers MJ, Nechemias C, Smith H. Infection and diabetes: the case for glucose control. Am J Med 1982; 72:439–50. Fig. 2. Right Comparison of incremental rise in plasma glucose at completion of single hormone infusion with that observed after combined infusion of three hormones. *Reproduced* with permission from Shamoon H, Hendler R, Sherwin RS. Synergistic interactions among antiinsulin hormones in the pathogenesis of stress hyperglycemia in humans. J Clin Endocrinol Metab 1981; 52:1235–41.

Weddell and Gale published observations on the effect of surgery on glucose level in patients without diabetes who did not receive glucose perioperatively (9). The researchers reported that the majority of their patients experienced a rise in ambient glucose level, many of the levels exceeding 200 mg/dL. This perioperative rise in glucose in normal individuals, as well as the difficulty in controlling diabetes during surgery, has been largely ascribed to hormonal changes.

In this section I evaluate the various hormones altered by surgery and discuss how they work synergistically to effect a perioperative rise in the glucose level. In particular, I focus attention on the effect of surgery on the "counterregulatory hormones": glucagon, catecholamines, growth hormone, and cortisol. These are the hormones that the body releases during hypoglycemia in order to raise the plasma glucose.

**Glucagon.** In 1975 Russell et al. evaluated the effect of surgery on glucagon levels (10). Glucagon is a hormone that functions to release glucose stored in the liver as glycogen, and therefore raises the plasma glucose. Russell et al. measured glucagon levels in 21 nondiabetic patients who underwent a variety of surgical procedures. Measurements for glucagon were performed at least twice prior to surgery, every 10–15 min throughout surgery, and then daily for the first seven postoperative days. Glucagon levels were found to be stable preoperatively, but increased significantly during the operative period, and remained significantly elevated through postoperative day 4. The plasma glucose level also rose during surgery and was significantly elevated compared to basal levels. Consequently, surgery and the first couple of postoperative days are associated with an elevation in the glucose-raising hormone glucagon.

**Catecholamines.** The effect of surgery on catecholamines was evaluated by Halter et al. in 1977 (11). Eight patients who underwent elective abdominal surgery had basal, intraoperative, and postoperative measurements of both norepinephrine and epinephrine. The results were compared to those of a control group of 11 healthy individuals. Halter et al. found a marked intraoperative and postoperative elevation in both catecholamines over basal preoperative levels as well as over levels in the control groups. The authors concluded that "the changes of plasma levels of NE (norepinephrine) and E (epinephrine) observed in this study provide direct evidence of adrenergic activation during elective abdominal surgery in man."

**ACTH and Growth Hormone.** The impact of surgery on the counterregulatory hormones ACTH and growth hormone was investigated by Newsome and Rose in 1971 (12), who documented that patients undergoing general anesthesia for either a cholecystectomy or inguinal herniorrhaphy had a marked intraoperative rise in ACTH.

Furthermore, the study found an increase in human growth hormone in men undergoing inguinal herniorrhaphy but not in obese women who had cholecystectomies.

Thus research has clearly shown that the stress of surgery is associated with secretion of hormones typically released by the body in defense against hypoglycemia.

**Hormonal Increases and Intraoperative Glucose Levels.** The marked elevation of the counterregulatory hormones during surgery raises the question of impact on intraoperative glucose levels. To address this question, the effect of counterregulatory hormones in normal individuals was evaluated by Shamoon et al. in 1981 (13) in a study specifically designed to determine the impact of the release of the counterregulatory hormones (cortisol, glucagon, and epinephrine) on a normal nondiabetic individual. Each of the hormones was infused individually after an overnight fast in doses typically found during severe stress; glycemic response was then evaluated. Following the separate administrations, all three hormones were infused simultaneously. Shamoon et al. found that separate administration of the hormones elicited a minor increase in the ambient glucose level, but simultaneous delivery of all three counterregulatory hormones caused a marked increase in the plasma glucose to levels over 200 mg/dL (Fig. 2). The authors concluded, "Our data suggest that stress-induced hyperglycemia cannot be ascribed to a single antiinsulin hormone, but, rather, may be a consequence of synergistic hormone interactions." An increase in the glucose level to values exceeding 200 mg/dL was exactly what Weddell and Gale found in their 1934 article on the response of nondiabetics to surgery (9).

**Paradoxic Intraoperative Response of Insulin.** Adding to the metabolic turmoil of surgery is the paradoxical response of insulin during the intraoperative period. Russell et al. demonstrated that despite an increase in glucose during surgery, a significant decrease in plasma insulin level occurred (10). Halter and Pflug confirmed this finding, showing a 50% reduction in insulin levels during anesthesia over preanesthesia levels in 16 nondiabetic men having elective surgery (14). The study concluded that "minor surgical procedures performed with inhalation anesthesia resulted in hyperglycemia and impaired insulin responses to IV (intravenous) glucose."

**Optimal Environment for Hyperglycemia.** Given the decreased secretion of insulin and the outpouring of counterregulatory hormones, an optimal environment is established for hyperglycemia during surgery. This manifests itself in hyperglycemia in nondiabetic individuals as well as creating the potential for marked hyperglycemia in individuals with diabetes. In fact, the hormonal changes that occur during surgery, namely the elevated glucagon levels with concomitant insulin suppression, are the classical changes found in patients with type I diabetes experiencing diabetic ketoacidosis (15). On a more practical level, most diabetologists have seen type I patients go into ketoacidosis and type II patients develop a hyperosmolar hyperglycemic state in the postoperative period. Clearly, maintaining glucose homeostasis during surgery and the postoperative period is an important yet formidable task.

## Practical Guidelines for Perioperative Glucose Control

The goal of the physician managing the care of the diabetic in the perioperative period is to achieve a state of glycemic control that protects against the dangers of hyperglycemia (intraoperative metabolic abnormalities, postoperative infections, and decreased integrity of the wound) without producing severe and potentially life-threatening hypoglycemia. To accomplish this difficult task at a time when counterregulatory hormones are high requires a coordinated effort between the diabetologist, surgeon, anesthesiologist, and recovery room staff, as well as the staff nurses on the floor. It is crucial that this effort be coordinated so that the desired outcome is achieved.

**Desirable Level: 150–250 mg/dL.** Of utmost importance is a clear definition of the plasma glucose level that is desired. In the literature, a glucose level greater than 250 mg/dL has been associated with poor wound healing. However, it is unclear if lowering the glucose level below this point will improve wound integrity. Furthermore, the optimal glucose level needed to decrease postoperative wound infections is not known, but levels exceeding 250 mg/dL clearly have a deleterious effect on in vitro functioning of white blood cells. Therefore, I recommend keeping the perioperative plasma glucose level between 150 and 250 mg/dL, a level that represents a compromise between minimizing the risk of hyperglycemia and protecting against hypoglycemia.

**General Principles.** Five general principles can be applied to patients with diabetes who are having surgery; adherence to these principles will help optimize perioperative glucose control:

- **Monitor glucose frequently.** Perhaps the most important aspect of perioperative glucose control is frequent glucose monitoring. This allows for close regulation of the glucose level and alerts the physician that corrective action needs to be taken. Without recurrent glucose analysis, perioperative glucose control is unattainable, and the potential for marked hypoglycemia or hyperglycemia is great. Close monitoring must continue until the patient is able to eat.
- **Optimize glucose control preoperatively.** Fletcher et al. documented in 1965 that the major determinant of intraoperative glucose levels was preoperative control (16). Consequently, the better control the patient has preoperatively, the easier it will be to effect good perioperative control.
- **Clarify the type of diabetes prior to surgery.** In order to deliver appropriate perioperative management it is crucial that the physician determine whether the patient is a type I diabetic (insulin-dependent, ketosis-prone). This has obvious implications for insulin use intraoperatively and will guide the physician's perioperative management.
- **Schedule all elective cases for early in the morning.** Whenever feasible, a diabetic who is having surgery should be scheduled as the first case of the morning. There is a direct correlation between the number of hours a diabetic is not allowed to eat and the chance of marked hypoglycemia or hyperglycemia. Consequently, delaying the surgery until later in the day increases the likelihood of metabolic complications.
- **Perform a thorough cardiovascular examination preoperatively.** Both men and women with diabetes are known to be at increased risk for atherosclerotic heart disease. Furthermore, angina and myocardial infarction may be silent owing to the autonomic neuropathy of diabetes. Therefore, it is crucial that the cardiac status of the patient be evaluated prior to surgery. This recommendation has been highlighted by a recent study by MacKenzie and Charlson (17), a retrospective review of all diabetics who underwent surgery at New York Hospital in 1980 that found a preoperative history of congestive heart failure or valvular heart disease to be predictive of perioperative cardiac morbidity and mortality.

**Type I Diabetes.** Perioperative management of the patient with type I diabetes is straightforward. Owing to the absence of endogenous insulin, the type I diabetic requires exogenous insulin at all times. In the absence of insulin, the patient will become catabolic and develop simultaneous proteolysis and lipolysis. Within 24 hr, the metabolic state will decompensate into diabetic ketoacidosis. Consequently, insulin must be administered throughout the perioperative period, even though the patient is not yet allowed to eat. Therefore, to protect against the hypoglycemic effect of insulin, intravenous dextrose must be administered.

Of the variety of regimens for managing the type I diabetic, the safest is probably the following:

- Give the usual regimen of insulin the day prior to surgery.
- At 6 AM the day of surgery begin an intravenous drip of D5 1/2 NS at 100 mL/hr. Added to this drip should be 10 units of regular insulin. The patient will receive 5 g of carbohydrate an hour and one unit of insulin.
- Blood glucose determination should be performed every 1–2 hr. If the glucose level exceeds 250 mg/dL, more regular insulin should be added to the intravenous solution. If the glucose falls below 100 mg/dL, a new drip with less insulin must be instituted, or more dextrose must be given (possibly as D50).

The advantage of this protocol is that at all times insulin and dextrose are administered simultaneously and the patient is protected against infusion of either separately. It is critical that the combined insulin and glucose infusion be continued until the patient has returned to his or her regular diet.

**Type II Diabetes.** Perioperative management of the patient with type II diabetes is more varied:

- If the patient is managed preoperatively on diet therapy or oral hypoglycemic agents, the perioperative management consists of frequent glucose monitoring and regular insulin administration as needed.
- The type II diabetic who requires multiple insulin injections to achieve glucose control as an outpatient should be placed on the combined insulin/dextrose protocol described for patients with type I diabetes.
- Controversy still exists on appropriate perioperative management of the type II diabetic who has been taking a single injection of insulin daily. In this circumstance, clinical acumen is paramount. The two alternatives are the insulin/glucose protocol or the administration of approximately half the usual morning dose of insulin along with an infusion of D5W at 100 ml/hr. Because glycemic control is easier to achieve with

the combined insulin/dextrose infusion, this protocol should be used in those patients who are markedly hyperglycemic preoperatively or are undergoing major surgery.

## Summary

Perioperative management of the patient with diabetes presents unique challenges. The benefit of glycemic control must be balanced against the danger of hypoglycemia. Through a clear understanding of the hormonal and metabolic alterations that accompany surgery, as well as careful glucose monitoring, the diabetic can be safely and effectively managed during the perioperative period.

## References

1. McMurry JF. Wound healing with diabetes mellitus: better glucose control for better wound healing in diabetes. Surg Clin North Am 1984; 64:769–778.
2. Goodson WH, Hunt TK. Wound healing and the diabetic patient. Surg Gynecol Obstet 1979; 149:600–608.
3. Goodson WH, Hunt TK. Studies of wound healing in experimental diabetes mellitus. J Surg Res 1977; 22:221–227.
4. Gottrup F, Andreassen TT. Healing of incisional wounds in stomach and duodenum: the influence of experimental diabetes. J Surg Res 1981; 31:61–68.
5. Mowat AG, Baum J. Chemotaxis of polymorphonuclear leukocytes from patients with diabetes mellitus. N Engl J Med 1971; 284:621–627.
6. Bagdade JD, Nielson KL, Bulger RJ. Reversible abnormalities in phagocytic function in poorly controlled diabetic patients. Am J Med Sci 1972; 263:451–456.
7. Cruse RJE, Foord R. A five-year prospective study of 23,649 surgical wounds. Arch Surg 1973; 107:206–210.
8. Rayfield EJ, Ault MJ, Keusch GT, Brothers MJ, Nechemias C, Smith H. Infection and diabetes: the case for glucose control. Am J Med 1982; 72:439–450.
9. Weddell AG, Gale HED. Changes in the blood-sugar level associated with surgical operations. Br J Surg 1934; 22:80–87.
10. Russell RCG, Walker CJ, Bloom SR. Hyperglucagonemia in the surgical patient. Br Med J 1975; 1:10–12.
11. Halter JB, Pflug E, Porte D. Mechanism of plasma catecholamine increases during surgical stress in man. J Clin Endocrinol Metab 1977; 45:936–944.
12. Newsome HH, Rose JC. The response of human adrenocorticotrophic hormone and growth hormone to surgical stress. J Clin Endocrinol Metab 1971; 33:481–487.
13. Shamoon H, Hendler R, Sherwin RS. Synergistic interactions among antiinsulin hormones in the pathogenesis of stress hyperglycemia in humans. J Clin Endocrinol Metab 1981; 52:1235–1241.
14. Halter JB, Pflug E. Relationship of impaired insulin secretion during surgical stress to anesthesia and catecholamine release. J Clin Endocrinol Metab 1980; 51:1093–1098.
15. Stagnaro-Green AS. Diabetic ketoacidosis—in search of zero mortality. Mt Sinai J Med 1990; 57:3–8.
16. Fletcher J, Langman MJS, Kellock TD. Effect of surgery on blood sugar levels in diabetes mellitus. Lancet; 1965; 2:52–54.
17. MacKenzie CR, Charlson ME. Assessment of perioperative risk in the patient with diabetes mellitus. Surg Gynecol Obstet 1988; 167:293–299.

# Intensive Insulin Therapy:

## Euglycemia with a Price Tag

SHEILA H. ROMAN, M.D.

IN THE late 1970s new tools for the assessment and management of individuals with diabetes became widely available. Technology allowing self-monitoring of blood glucose levels permitted patients to be educated to make their own daily insulin decisions at home. The glycosylated hemoglobin assay provided a reliable quantitative assessment of the "average" glycemic levels over the preceding 6–12 weeks. In the 1980s a number of investigators (1–5) demonstrated that carefully selected patients could be trained to combine self-monitoring of blood glucose levels with insulin infusion devices or multiple insulin injections in complex regimens designed to maintain glycosylated hemoglobin values close to the normal range. Thus the era of intensive insulin therapy for diabetes management was begun.

Intensive insulin therapy can be defined as the use of multiple (3 or more) daily insulin injections or use of an insulin delivery device such as a continuous subcutaneous insulin infusion pump to maintain near-normoglycemia. I review here the rationale for intensive insulin therapy as well as the indications and contraindications for its use. Insulin actions, target blood sugars, and a variety of insulin regimens are presented, including progression from minimal to average to intensive levels of insulin treatment. The complications and risks of intensive insulin therapy are highlighted, and guidelines are proposed to maintain minimum risk at maximum benefit.

### Rationale

The goal of intensive insulin therapy in diabetes management is to approximate nondiabetic physiology as closely as possible and bring blood glucose levels close to the normal range pre- and post-prandially in diabetic individuals. The rationale for using intensive insulin therapy to lower blood glucose levels to the normal range is the prevention of diabetic complications. Although many studies have provided evidence for an association between the duration and severity of hyperglycemia and the development of diabetic microvascular disease and neuropathy (6, 7), maintaining euglycemia is still not proven by rigorous clinical trials to delay or prevent diabetic complications. In experimental animal models of diabetes, background retinopathy and nephropathy were prevented or ameliorated with treatment measures that reduced hyperglycemia (8, 9). In humans many studies involving relatively short periods of improved diabetic control have resulted in improved blood lipid levels, skeletal growth, nerve conduction velocities, renal filtration, microalbuminuria, and muscle capillary basement width (10–15). However, most human studies have been either cross-sectional or nonrandomized follow-up studies of an epidemiologic nature. Few studies were randomized, controlled clinical trials, and most studies involved relatively small numbers of patients.

Proof that a causal relationship exists between a reduction in the degree of hyperglycemia and a reduction in the risk of diabetic complications requires a prospective controlled study in-

---

Adapted from the author's presentation at the symposium "Diabetes Mellitus: Strategies for the 1990s" held at the Mount Sinai Medical Center on April 24, 1990.

From the Adult Clinical Diabetes Program, Department of Medicine, Mount Sinai School of Medicine (CUNY), New York, New York.

Address reprint requests to Sheila H. Roman, M.D., Chief, Adult Clinical Diabetes Program, Department of Medicine, Mount Sinai School of Medicine (CUNY), One Gustave L. Levy Place, New York, NY 10029.

TABLE 1
*Insulin Action*

| Insulin | Onset (hr) | Peak (hr) | Usual effective duration |
|---|---|---|---|
| Animal | | | |
| Regular | 0.5–2.0 | 3–4 | 4–6 |
| NPH | 4–6 | 8–14 | 16–20 |
| Lente | 4–6 | 8–14 | 16–20 |
| Ultra Lente | 8–14 | Minimal | 24–36 |
| Human | | | |
| Regular | 0.5–1.0 | 2–3 | 3–6 |
| NPH | 2–4 | 4–10 | 10–16 |
| Lente | 3–4 | 4–12 | 12–18 |
| Ultralente | 6–12 | Minimal | 18–20 |

*Adapted* with permission from a slide made available by the American Diabetes Association.

TABLE 2
*Target Blood Glucose Levels*
*(mg/dL)*

| | Ideal | Acceptable | Safest |
|---|---|---|---|
| Before breakfast | 70–90 | 70–110 | 90–130 |
| Before lunch, supper, and bedtime snack | 70–105 | 70–130 | 90–150 |
| 1 hr after meals | <160 | <180 | <200 |
| 2 hr after meals | <120 | <150 | <170 |
| 2 A.M.–4 A.M. | >70 | >70 | >90 |

volving a large number of patients. In 1982 the National Institute of Arthritis, Diabetes, and Digestive and Kidney Diseases initiated the Diabetes Control and Complications Trial, a multicenter, randomized clinical trial to prospectively study the effects of a standard (up to two dose) insulin treatment regimen and an "experimental" (multiple daily insulin injections or continuous subcutaneous insulin infusion) treatment regimen on the development or progression of early vascular complications of people with type 1 diabetes. Results from this trial should start to become available in 1993 and should definitively answer the question of whether lowering blood glucose levels to near-normal values with intensive insulin therapy prevents microvascular complications.

Until the results of the study are available, the burden of proof is on those who maintain that diabetes control is without effect on the complications of diabetes. Therefore, most diabetologists strongly believe that the goal of appropriate diabetes management should include a serious effort to achieve levels of blood glucose as close to those of a nondiabetic person as feasible.

## Indications and Contraindications

Intensive insulin therapy is indicated for:

- diabetics in otherwise good health and without hypoglycemic risk factors or advanced diabetic complications
- pregnant diabetic mothers beginning before conception
- diabetics following renal transplantation

A large number of such patients will be adolescent or young adults with type 1 diabetes.

There are many contraindications to intensive insulin therapy. One of the primary contraindicators is life expectancy limited by advanced age or other life-shortening disease. Since the aim is prevention of complications that require 5 to 20 years to develop, patients with a life expectancy less than approximately 5–10 years are not candidates.

Patients with advanced complications of diabetes should also be excluded because there is no scientific basis for the reversal of end-stage microangiopathic lesions by meticulous blood sugar control. Patients with advanced diabetic complications are usually also at high risk for severe hypoglycemia.

Since the major risk of intensive insulin regimens is hypoglycemia, such regimens should not be prescribed for patients for whom hypoglycemia poses a high risk. In particular, patients with significant cerebrovascular or coronary artery disease should be rejected because of the potentially lethal effect of hypoglycemia.

Severe autonomic neuropathy would also be a contraindication to intensive insulin therapy because of the loss of the catecholamine response to hypoglycemia in such patients. A large number of patients who have had diabetes for more than five years have some loss of catecholamine response to hypoglycemia, and therefore "hypoglycemia unawareness" in general is a relative contraindication to intensive insulin regimens.

Patients with moderate to severe degrees of renal insufficiency have prolonged and often unpredictable insulin action and generally are not candidates for this therapy.

Patients with liver failure should not receive intensive insulin treatment.

Individuals with psychiatric disturbances, alcohol or drug habits, and others who do not exhibit the required level of reliability and intelligence to understand and implement the principles of diabetic self-care should also be excluded.

### Intensive Insulin Management

Review of the various insulin preparations and their action is required for any discussion of intensive insulin regimens. Three types of insulin—rapid-acting (regular), intermediate-acting (NPH and lente), and long-acting (ultralente)—differ in onset, duration, and peak of insulin action (Table 1). The human insulin preparations in wide use today all have shorter durations of action and, in a number of cases, earlier onsets and peaks than the animal insulins previously used.

The objective of intensive insulin therapy is to maintain blood glucose levels as close as possible to a normal range throughout a 24-hr period. The "ideal" targets are defined by the glucose levels in nondiabetic subjects; the recommended "safest" levels result in glucose values greater than those of nondiabetic subjects throughout a 24-hr period, but still result in good glycemic control and minimize the risk of hypoglycemia substantially (Table 2).

**Levels of Insulin Treatment.** A minimal level of insulin treatment resulting in poor metabolic control would include the following biochemical and clinical characteristics: hemoglobin A1C of >12% (normal 4.5%–6.6%), many blood glucose values greater than 300, almost constant glucosuria, and intermittent spontaneous ketonuria. An example of a minimal level of treatment would be the prescription for a type 1 diabetic patient of one daily injection of insulin. Such a level of treatment is inappropriate for any patient.

An average level of insulin treatment would include the following biochemical and clinical characteristics: hemoglobin A1C of 8%–10% (normal 4.5%–6.6%), premeal blood glucose values of 150–200 mg/dL, intermittent glucosuria, and rare ketonuria. An example of an average level of treatment would be a two-dose insulin regimen with either NPH or mixed NPH and regular insulin before breakfast and before dinner. This level of treatment may be appropriate for many clinical circumstances.

Intensive levels of insulin treatment include the following biochemical and clinical characteristics: hemoglobin A1C of 6%–7% (normal 4.5%–6.6%), premeal blood glucoses between 70 and 120 mg/dL, and postmeal self-monitored blood glucose of less than 180 mg/dL. There should essentially be no positive urine glucoses or ketones. To achieve such blood glucose levels, the regimen must supply both basal and increased insulin action for meals and snacks. Such regimens must be able to take into account the dawn phenomenon, the physiologic tendency for blood glucose to drift upward at dawn secondary to sleep-induced growth hormone release. Either multiple-dose insulin regimens or subcutaneous insulin infusion with an open-loop insulin pump can provide this kind of insulin action.

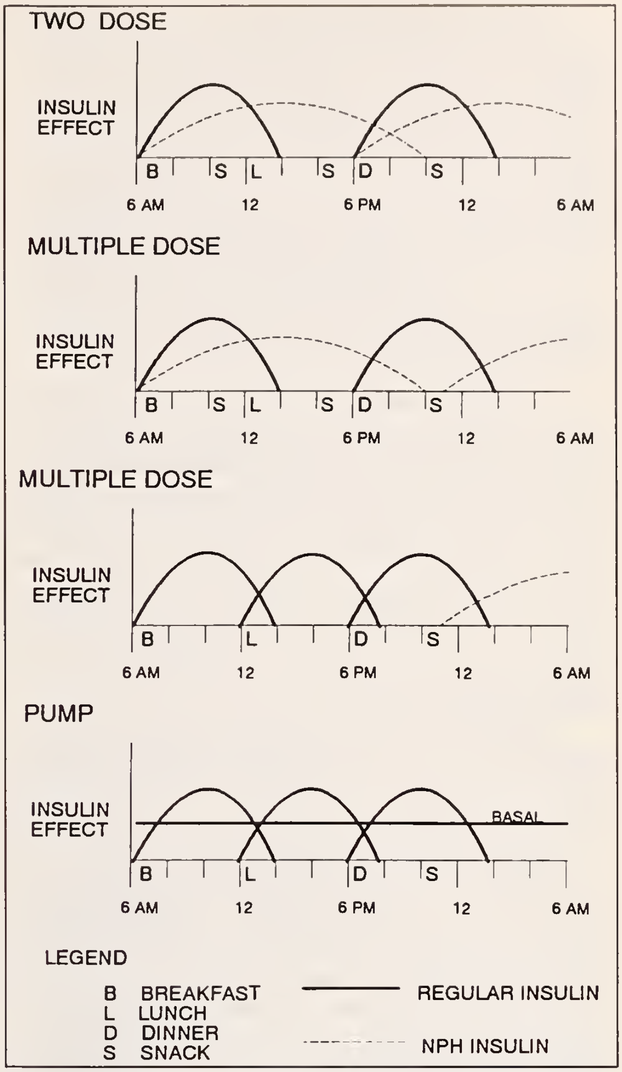


Fig. Examples of a variety of multiple-dose insulin delivery schemes.

**Multiple-Dose Regimens and Insulin Pumps.** Multiple-dose insulin regimens provide basal insulin action through a depot source of an intermediate (NPH or lente) or long-acting (ultralente) insulin and provide regular insulin action to cover meals and snacks. Generally the intermediate or long-acting form of insulin is used in a daily regimen to provide a basal rate of insulin release from a depot site, and the patient is then instructed to use a sliding scale of regular insulin for coverage of meals throughout the day. Additional intermediate-acting insulin may be given at bedtime, depending on the regimen, to cover the dawn phenomenon.

The number of regular insulin injections will vary depending on the complexity of the insulin regimen, the type of depot insulin used (NPH or lente versus ultralente), and the overlap of insulin actions throughout the day. Patients must be carefully instructed on the use of an algorithm for premeal regular insulin coverage and also as to the time of day when they will have multiple insulin actions overlapping and the risk of hypoglycemia will be greatest. Examples of a variety of multiple-dose insulin delivery schemes are shown in the Figure.

For certain patients, insulin pumps may provide greater life-style flexibility, in particular with regard to meal schedules and travel. Insulin pumps deliver regular insulin throughout the day and evening at a predetermined basal infusion rate. On demand from the user, the pump can give a bolus of regular insulin, to cover eating or snacking. The pumps can also be programmed to increase their basal infusion rate to cover the dawn phenomenon.

The use of an insulin pump is labor intensive for both patient and physician, since basal insulin rates and bolus doses must be specifically tailored to the individual pump user.

There are no well-controlled clinical trials that offer compelling evidence for the superiority of blood sugar control by either multiple insulin injections or insulin pump. The choice between multiple-dose and insulin pump therapy usually comes down to a personal preference of the physician. Since there is not a clear-cut glycemic advantage, my service tends to favor multiple-dose intensive insulin therapy.

**The Patient's Role in Intensive Regimens.** Patients who utilize an intensive insulin regimen must understand that they must take an active role in their diabetes care. Patients must have a realistic perspective about their diabetes management and thoroughly understand that these regimens will require even more attention to their diabetes. These regimens require testing blood sugar levels at least four times a day and learning how to make proper adjustments in insulin. In particular, insulin pump users must understand that the pump is not an artificial pancreas and requires intensive monitoring of blood glucose, frequently up to seven times a day. Also, because there is no depot dose of insulin available, pump patients must understand that they have no available insulin reserve without the pump. They must be able to respond quickly to pump malfunctions.

**Initiating Intensive Therapy.** To initiate an intensive insulin therapy regimen in a patient being treated by a standard insulin regimen, the first step is to conduct baseline evaluations, including glucose, hemoglobin A1C, lipids, and ophthalmologic and renal parameters. Then the patient is switched from a once- or twice-daily injection regimen to a split, mixed insulin regimen including short- and intermediate-acting insulin twice a day. Short-acting insulin would be adjusted to an individual algorithm. The patient should have nutritional evaluation and focus on this advice. Patients must monitor their capillary blood glucose levels four times a day and be instructed to identify all probable causes of hyper- or hypo-glycemia they encounter. Initial training in this technique is labor intensive for both the patient and health care provider.

The self-monitoring of blood glucose may need to be increased to include postprandial and nocturnal measurements in order to determine the patient's blood glucose pattern in relation to insulin action. The insulin regimen might then be changed from a two to a three or four times daily insulin regimen, depending on the patient's needs and blood glucose pattern. The depot insulin type might be changed from an intermediate-acting to a long-acting insulin. For selected patients, an insulin pump might be considered.

**Complications.** Early Diabetes Control and Complications Trial data document a higher frequency of both severe and mild hypoglycemic episodes among patients undergoing intensive insulin therapy than among those undergoing conventional insulin therapy (16). In the study the increase in both severe and mild hypoglycemia was twofold to threefold. Hypoglycemia remains the most frequent and serious complication of intensive insulin therapy.

Data from the first year of experience of the study also document a weight gain in type 1 diabetic patients on intensive insulin therapy. The clinical significance of the weight gain remains unclear. Several other consequences of intensive insulin therapy are particular to insulin-pump–treated patients, including increased incidence of diabetic ketoacidosis due to the lack of depot insulin, and local abscesses at the site of the indwelling needle.

**Guidelines.** Intensive insulin therapy is a feasible therapeutic option for patients with diabetes. However, it is labor intensive for both the patient and members of the health care team. Intensively controlled patients must be carefully selected and must understand fully all the techniques involved and the potential risks of the regimen.

Such understanding can only be accomplished through a team approach that includes physician, diabetes nurse specialist, and nutritionist. In particular, patients must be trained in how to prevent and treat hypoglycemia. This requires a period of intensive instruction by a certified diabetes educator in: insulin actions; timing of insulin, meals, and exercise; "high risk" times of the day; and early recognition and treatment of hypoglycemic symptoms.

Patients should be advised to wear a Medic Alert and always carry a standardized source of glucose (such as Glucose-Tabs or Insta-glucose).

Family members should be instructed on the administration of glucagon.

Patients must be willing to self-monitor their blood glucose levels at least four times a day. This daily self-monitoring program must be designed to assure the patient's safety on a day-to-day basis and may include testing during the middle of the night on occasion to evaluate for and reduce the risk of nocturnal hypoglycemia. Patients must be carefully trained to make appropriate precautionary adjustments in their insulin regimens to prevent hypoglycemia during sleep and during physical exercise.

The potential long-term benefits of meticulous glucose regulation must be carefully weighed against the short-term risks, particularly the risk of hypoglycemia.

## Summary

The benefit of lowering blood glucose levels with intensive insulin therapy, especially with regard to the development and progression of microvascular complications, is still being investigated. Although many studies have demonstrated a correlation between the degree of hyperglycemia and the risk of development of long-term diabetic complications, we still do not know at what level of glycemia that risk occurs. It is hoped the Diabetes Control and Complications Trial will answer the question. Until that time, patients who are candidates for intensive insulin regimen therapies should be educated and trained to regard the prevention of hypoglycemia as the highest priority of their self-care.

### Acknowledgments

This article is based, in part, on a lecture slide series made available by the American Diabetes Association.

### References

1. Pickup JC, et al. Long-term continuous subcutaneous insulin infusion in diabetics at home. Lancet 1979; 2:870–873.
2. Tamborlane WV, et al. Outpatient treatment of juvenile-onset diabetes with a preprogrammed portable subcutaneous insulin infusion system. Am J Med 1980; 68:190–196.
3. Schiffen A, Colle E, Belamonte M. Improved control in diabetes with continuous subcutaneous insulin infusion. Diabetes Care 1980; 3:643–649.
4. Champion MC, Shepherd GAA, Rodger NW, Dupre J. Continuous subcutaneous insulin infusion in the management of diabetes mellitus. Diabetes 1980; 29:206–212.
5. Skyler JS, Skyler DL, Seigler DE, O'Sullivan M. Algorithms for adjustment of insulin dosage by patients who monitor blood glucose. Diabetes Care 1981; 4:311–318.
6. Pirart J. Diabetes mellitus and its degenerative complications: a prospective study of 4400 patients observed between 1947 and 1973. Diabetes Care 1978; 1:168–188.
7. Tchobroutsky G. Relation of diabetic control to development of microvascular complications. Diabetologia 1978; 15:143–152.
8. Engerman R, Bloodworth JB, Nelson S. Relationship of microvascular disease in diabetes to metabolic control. Diabetes 1977; 26:760–769.
9. Mauer S, Steffes M, Brown D. The kidney in diabetes. Am J Med 1981; 70:603–612.
10. Tamborlane W, Genel M, Sherwin R, Felig P. Restoration of normal lipid and amino acid metabolism in diabetic patients treated with a portable insulin infusion pump. Lancet 1979; 1:1258–1261.
11. Tamborlane WV, Hintz RL, Bergman M, Benel M, Felig SP, Sherwin RS. Insulin-infusion pump treatment of diabetes: influence of improved metabolic control on plasma somatomedin levels. N Engl J Med 1981; 305:303–307.
12. Pietri A, Ehle A, Raskin P. Changes in nerve conduction velocity after six weeks of glucose regulation with portable insulin infusion pumps. Diabetes 1980; 29:668–671.
13. Viberti G, Keen H. The patterns of proteinuria in diabetes mellitus: relevance to pathogenesis and prevention of diabetic nephropathy. Diabetes 1984; 33:686–692.
14. The Kroc Collaborative Study Group. Blood glucose control and the evaluation of diabetic retinopathy and albuminuria. N Engl J Med 1984; 311:365–372.
15. Peterson CM, Jones RL, Esterly JA, Wantz GE, Jackson RL. Changes in basement membrane thickening and pulse volume comcomitant with improved glucose control and exercise in patients with insulin-dependent diabetes mellitus. Diabetes Care 1980; 3:586–589.
16. The DCCT Research Group. Diabetes Control and Complications Trial (DCCT): results of feasibility study. Diabetes Care 1987; 10:1–19.

*Grand*
*Rounds*
*In*
*Psychiatry*

# Interfaces between Seizures and Affective Disorders:

## The Uses of Visually Mapping the Evolution and Longitudinal Course of an Illness

ROBERT M. POST, M.D., ROBERT L. FINDLING, M.D., AND RENE S. KAHN, M.D.

**Rene S. Kahn:** The case presentation is from the Schizophrenia Research Unit at the Bronx Veterans Administration Medical Center.

I think we found, by pure coincidence, an interesting patient to illustrate the potential interfaces between seizures and affective disorders. We selected a patient who was admitted to the hospital to be studied as a schizophrenic patient, but he was not schizophrenic—at least we think he wasn't—and I think the case is an instructive one. Dr. Findling presents the case and Dr. Post then discusses it.

### Diagramming the Course of Illness

**Robert M. Post:** In keeping with trying to make this an informal case description, I am going to start off by making a diagram of the case presentation simultaneously with the presentation. How many of you actually use a visual graph of the course of illness when you're doing a consultation with a bipolar patient? [Few answer affirmatively.] I hope you will take the message as the main message of this case, that often something valuable is to be gained from developing a detailed life-chart (1–3). Many people assume that making a life-chart like this is hard and complicated; I'll show you that it really isn't. A life-chart can be constructed from a two-minute telephone conversation when your colleague calls you, or you can graph it from a more formal presentation, as we're going to do.

The life-chart process has many merits and brings into focus the entire longitudinal course of the patient's illness. Particularly with the affective illnesses, but also with schizophrenia, the focus should not always be on the acute symptom presentation, but also on where the patient has been, how the course of illness evolved, and where the patient is likely to go; these issues will help set up your long-term therapeutics. I think life-charting the patient longitudinally helps bring these principles into focus. We are going to talk about diagnosis, and about visually graphing patients' course of illness.

### Case Report

**Robert L. Findling:** The patient is a 50-year-old, right-handed, divorced, unemployed white man who was admitted to the Schizophrenia Research

---

Adapted from the authors' Grand Rounds in Psychiatry presentation at the Mount Sinai Medical Center January 30, 1990, which was funded in part by an educational grant from The Upjohn Company.

Case Presenters: Robert L. Findling and Rene S. Kahn; Discussant: Robert M. Post.

From the Biological Psychiatry Branch, National Institute of Mental Health, Bethesda, MD (RMP), and the Department of Psychiatry, The Mount Sinai Hospital, New York, NY (RSK, RLF). Address correspondence and reprint requests to Robert M. Post, M.D., Chief, Biological Psychiatry Branch, National Institute of Mental Health, Building 10, Room 3N212, 9000 Rockville Pike, Bethesda, MD 20892.

Unit at the Bronx Veterans Administration Hospital with a diagnosis of paranoid personality disorder on August 31, 1989.

**History of Present Illness.** The patient was well until three days after Memorial Day weekend in 1988, when he began acting "irrationally." He did not remember what he was doing at the time. The girlfriend with whom he lived stated that after Memorial Day weekend in 1988, he became "very upset." A few days after the holiday, the patient became irritable and subsequently started acting like a "big shot." He dressed in expensive suits, would walk down the middle of the street giving money away. The patient would sit under a chair and talk on the phone while no one was on the line. He would also talk in a loud voice as if others were in the room, and also felt that he was a target of a Mafia conspiracy. Sometimes he believed he was an FBI agent and called the police frequently to inform them of the whereabouts of certain spies. The patient denied having delusions of thought insertion, withdrawal, or being controlled, or other bizarre delusions. Although the girlfriend noticed a decreased need for sleep, she did not note hypersexuality or increased rate of speech in the patient.

She took him to Harrington Manor Hospital, where he was hospitalized for six weeks. At the time of discharge neither the patient nor his girlfriend was informed of his diagnosis, and neither remembers the medications with which he was treated.

One month after this discharge, the patient was rehospitalized, this time at Bergen Pines Hospital in New Jersey, from July 28, 1988 until August 31 the same year. Again, the patient did not remember what caused him to be readmitted, but according to his hospital discharge summary, his initial mental status examination noted that he believed he owned a $1.5 million corporation. He claimed he lost his job in May 1988 as a car salesman for selling too many cars. No looseness of associations was noted, and the patient denied homicidal and suicidal ideation and both auditory and visual hallucinations. He was described as suspicious, guarded, irritable, and argumentative.

The tentative diagnosis was delusional disorder, paranoid type. The patient was started on loxapine, benztropine, and 500 mg/day of thioridazine but was subsequently noted to be "confused." An EEG with low temporal leads was interpreted as normal. A CT scan with and without contrast was also performed and read as normal. His thioridazine was decreased to 250 mg/day, and a subsequent decrease in confusion was noted. As the patient's delusional symptoms dissipated, the patient was thought to have developed "affective symptoms." He was noted to be irritable and brusque, with episodes of being inappropriately cheerful. Thyroid function tests were checked and were normal. The loxapine was discontinued and lithium carbonate was initiated, and a decrease in these "affective symptoms" was noted. He was discharged on thioridazine (250 mg/day) and lithium carbonate (300 mg three times a day), with follow-up at the Bergen Regional Mental Health Center.

His discharge diagnosis was delusional disorder, persecutory type.

On discharge the patient had plans of returning to car sales. Not feeling he needed medications on discharge from Bergen Pines, he stopped taking them and returned to live with his girlfriend in Philadelphia. For the next nine months, the patient was "unable to get out of bed." He had decreased energy and concentration and early and middle insomnia with a depressed mood. He frequently thought about loss of contact with his daughters and felt "guilty" about this. He felt unmotivated and incapable of holding a job. He was financially supported by his girlfriend during this time. The patient denied any weight loss or suicidal ideation during these months.

In the spring of 1989, one of the girlfriend's sons became ill. To pay for his health-care bills, she sold several pieces of furniture, including some that had originally belonged to the patient. The patient, seeing some of his possessions sold, felt that this woman was no longer worthy of his trust and decided to leave this woman with whom he had had a steady five-year relationship. He boarded a bus for Newark and on arrival there, telephoned his sister to ask if she could put him up temporarily. According to the patient, the sister told him she was scared of him and his erratic behavior and told him to leave her alone.

With nowhere to go, the patient took up residence in a men's shelter on Wards Island. About eight weeks into his stay at the shelter, he became more suspicious, distrustful, and anxious about people. He felt he could not trust anyone and said that he felt "paranoid." The patient also developed command auditory hallucinations of neutral content. The hallucinations were in an unknown voice and told the patient to do such things as get up, sit down, go to bed. The patient complained of difficulties initiating and maintaining sleep. He stated that these complaints brought him to seek psychiatric care. Of note, the patient also complained of what he called "time-losing spells." Again, the patient denied any bizarre delusions, delusions of being controlled, and hearing voices commenting about his actions.

He was referred to the Bronx Veterans Administration Medical Center, Psychiatric Evaluation Unit, from the men's shelter. The diagnosis by the Psychiatric Evaluation Unit was paranoid personality disorder. The patient was admitted to the Schizophrenia Research Unit as a potential research candidate.

**Medical History.** The medical history is remarkable for a crush injury to his left leg, which he acquired in the Navy when some machinery fell on him in 1958. In the mid 1980s, the patient had reconstructive surgery performed on the left leg. His only current sequela from this injury is that a pin had to be placed in his left ankle in 1958, which was still in place on admission. The patient walked with a cane and a limp due to intermittent ankle pain.

Otherwise remarkable in this patient's past medical history were the "time-losing spells." These episodes began in 1980 and were characterized as staring into space and being oblivious to his surroundings for several moments. They were associated with light-headedness and dimmed vision. The patient noticed he occasionally "blacked out." Witnesses stated that these attacks lasted approximately 30 seconds. Following these episodes, the patient would wake up and be confused for about five minutes. He could not remember what had happened during these times. The patient would occasionally "jerk" for several seconds during the episodes, as if startled. These episodes were not associated with an aura. The patient denied childhood seizures or a history of head trauma. These episodes occurred with a frequency of four times per year in 1980 and slowly increased in frequency, to once per month, then further increased; by the time of admission they were occurring up to once every other day.

The patient sought medical attention for these events in 1988, when the episodes became more frequent. He was twice seen in the Philadelphia VA emergency department for these complaints. Both times he was diagnosed as having "vasovagal episodes" and sent home.

**Psychiatric History.** Before his admissions in 1988, the patient had only one prior psychiatric contact. He was unable to give full details of this hospitalization but stated that in the early 1980s he was treated for about two months at Marlboro State Hospital in New Jersey. The patient was unsure why he was hospitalized, was not told his diagnosis, and does not remember the medications he received there.

When the psychiatric records from Marlboro State were obtained, it was found he was treated there for nine days in 1984 and given the diagnosis of schizophrenia. Two days prior to admission, the patient was fired from his job for being argumentative. On the day of admission he was found by the police lying on the railroad tracks stating that he was "just sunbathing." The admission note described the patient as angry, elated, and excited, with a labile affect. The note described the patient as coherent, related, but disorganized, with paranoid ideation. Again, no delusions of being controlled and no bizarre delusions or third-person auditory hallucinations were present. The patient was treated with 10 mg of haloperidol every four hours for a total of 18 consecutive doses; benztropine was given as needed. No discharge mental status examination was available, but the patient's ultimate prognosis was described as poor.

**Family History.** The patient has a 78-year-old mother in good health and a father who died from coronary artery disease. He has twin sisters aged 43 years, both of whom are healthy; he has two adopted daughters who are 25 and 19 years old. There is no family psychiatric history, except for a maternal sister who required institutionalization when elderly. Her diagnosis is not known. There is no family history of seizures.

**Social History.** The patient grew up in Hoboken, New Jersey, and stated that he had a happy childhood. He finished high school with OK grades, was able to keep part-time jobs during this time, joined the U.S. Navy in 1957, and served in the armed forces until 1963. He functioned as a gunner and during this time received the crush injury to his left leg, for which he underwent two reconstructive surgical procedures. For this injury the patient received a 10% disability status.

The patient was married in 1962. He and his ex-wife adopted two children and did so because he was told he was "sterile." After leaving the Navy in 1962–63, the patient became a car salesman, an occupation he maintained until Memorial Day weekend, 1988. The patient worked for numerous car firms, but according to him was quite successful. Indeed, at one point in his career, he was one of the 32 best Chrysler salesmen in the entire U.S. Just prior to admission at Willowbrook Manor, the patient was working 14-hour days and earning up to $2000 per week.

In 1979, the patient divorced his wife. According to the patient, the reason the divorce occurred was that he was a "workaholic" and never home. His frequent absences caused him to "grow apart" from his wife, prompting the divorce. He felt the divorce caused his daughters to think less of him and believed that his life took a downhill course somewhere between 1979 and 1980. He became more difficult to get along with and had to change jobs frequently due to "disagreements." He believed that his irritable behavior also contributed to his poor relationship with family members.

In 1984, while living alone near Philadelphia, the patient became involved in a single, steady relationship with a woman with whom he lived for five years. He moved out in July 1989. He stated he always had several friends—in fact, he states that he is still friends with many of the people with whom he grew up.

The patient stated that after his hospitalizations in 1988, his entire family wanted "nothing to do with him" because of his behavior. Since that time, he had had no contact with any family members, except for calling his sister from the Newark bus station in 1989. He said he was most upset by the loss of contact with his daughters. The patient smoked one pack of cigarettes a day for 30 years, denying alcohol and street-drug abuse.

Physical and neurologic examination results were within normal limits except for slight swelling and minor decreased range of motion of the left ankle. Results of the neurologic examination were unremarkable, with symmetric reflexes, no clonus, no Babinski sign, no frontal release sign, no ataxia, no weakness, no sensory deficit, and normal cranial nerve examination.

**Mental Status Examination on Admission.** The patient is a neatly dressed, well-groomed, slightly overweight man who looks his stated age. He walks with a limp and uses a cane, but ambulates without great difficulty. He is alert, attentive, and cooperative. His speech is quiet and clear, with normal cadence and rhythm. The patient is oriented to person, place, day, and date. His thought processes are goal-directed, coherent, logical, without flight of ideas or looseness of associations. His thought content is remarkable for the auditory hallucinations, but no visual hallucinations are present. He denies homicidal or suicidal ideation. He feels "paranoid," meaning he is afraid of people and believes that they want to hurt him. He feels "scared," which appears to be free-floating anxiety. He complains of problems in falling asleep and maintaining sleep. His mood is depressed and somewhat anxious. His affect is constricted. He denies helplessness, hopelessness, anergia, or anhedonia. He has difficulty concentrating due to his preoccupation with loss of contact with his family in general, but especially his daughters. He feels "awful" that his two adopted daughters do not wish to see him and sad because he lost his job and all his possessions. His judgment and insight are good. His short-term and long-term memory are intact. Finally, the patient also complains of being bothered by the "time-losing spells."

**Laboratory Results on Admission.** The patient had normal results on complete blood count and urinalysis; SMA-18 results were remarkable only for mildly elevated alkaline phosphatase (144 u/c). Results of his chest x-ray, EKG, thyroid levels, and coagulation studies were all normal. His PPD and VDRL results were negative.

**Summary.** This is a 50-year-old man who until 15 months prior to admission was functioning relatively well and had numerous stable relations both within and outside his family. After his hospitalizations in 1988, the patient lost his job and his friends and lost contact with all family members. His illness is characterized by periods of grandiosity, irritability, decreased sleep, and hallucinations; these periodic characteristics were subsequently followed by a single, more prolonged period of depressed mood, anergia, poor concentration, and early and middle insomnia, as well as feelings of guilt. He was referred to the Bronx Veterans Administration Psychiatric Evaluation Unit from a men's shelter with complaints of paranoia, command auditory hallucinations, insomnia, and "time-losing spells." He was subsequently admitted to the Schizophrenia Research Unit as a possible research candidate.

## Differential Diagnosis

**Findling:** Admission diagnoses include:

Axis I bipolar disorder depressed, rule out organic mood disorder.

Axis II, rule out paranoid personality disorder.

Axis III, rule out seizure disorder and status post crush injury to left leg.

**Post:** Let me stop here for a second. What have we got? (See Fig.) Anybody want to make a diagnosis?

**Audience:** Schizoaffective disorder.

**Post:** One issue is whether he has a schizoaffective diagnosis, and there's a question of whether he has any of these symptoms in his well intervals when he's not affectively ill. We haven't heard about that yet, but that perhaps will get clarified in the rest of the discussion. So we can leave that question open.

Are these spells absence seizures? If they are, we'd expect some spike wave in his EEG during them. What other possibilities do we have? Organic affective disorder? Manic-depressive illness due to underlying seizure disorder? The question is, if it's organic affective illness, how are the seizures and affective episodes related, since they are not closely related in time? We'll come back to that, but that's a key issue—the sequencing. Any other diagnoses?

Okay, so we've got a dual diagnosis with progression in both the seizurelike disorder, whatever it is, and an affective disorder.

**Bipolar Illness?** What else do we have for consideration in differential diagnosis? Nobody likes the schizophrenia diagnosis given in 1984. So everybody's agreed that although he looked schizophrenic on that admission (which he did), and on some of the other admissions, bipolar illness can present similarly and have just as many Schneiderian positive symptoms (4, 5). No one's bothered by the command hallucinations? O.K. They do not bother me either.

To make it explicit, we should state that typical, classical manic patients can be just as schizophrenic-looking cross-sectionally; what helps a lot is the fact that they function well premorbidly and intermorbidly. They have well intervals between episodes, so they are not schizophrenic even though cross-sectionally they look schizophrenic, and the differential diagnosis is essentially impossible to make, relying solely on an acute episode. So we discard the schizophrenia diagnosis—that's easy.

**Drug Use?** What else do we need to consider in differential diagnosis? Drug use? That's a good possibility to consider, because things like cocaine could produce seizures and affective episodes, but drug use doesn't seem to be part of the history.

**Childhood Abuse and Multiple Personality Disorder?** What else? What happens in the beginning of his life when he is only about five or seven years old? What possibilities do we have? No fevers, no seizures, we heard. This gentleman could still have a history of physical or sexual abuse as a youth and present with a pseudo-affective syndrome, something that looks just like classical affective illness and these weird, seizurelike spells; one really has to think about early childhood sexual or physical trauma with a presentation that can look very much like what we've heard now, with or without multiple personality disorder. It's something to keep one's eyes open for, particularly with these atypical spells, and, particularly, command auditory hallucinations. The people who have been traumatized often hear the voices of the people who are inflicting the physical trauma in a disguised fashion. Patients may not initially recognize the voices as those of alternate personalities, so with the history of these voices, we want to hear about his early development; as far as we know, it is negative for abuse. But that doesn't matter, because although some patients will offer a history of abuse to you as they walk into the office, other patients won't volunteer such information until they've been in treatment for several years. Even then they might not, and it might have to come out under hypnosis or an amytal interview or direct questioning. So if we eventually discard the seizure disorder as a real phenomenon, then we will need potential etiologies for these mysterious spells, and we should keep multiple personality disorder active in the differential diagnosis.

**Findling:** The question has been raised whether this man had fuguelike states, because there were some times when he was confused and did not always know what he was doing. Although the patient was acting in a bizarre fashion with poor memory during some of these events, they were not fugue states.

As for the nature of these "spells," these "time-losing spells" were similar. They were associated with several moments of not being aware of his surroundings, occasionally associated with all four extremities jerking. No tonic phase, but clonus in all four extremities. There were no hallucinations or incontinence during these episodes.

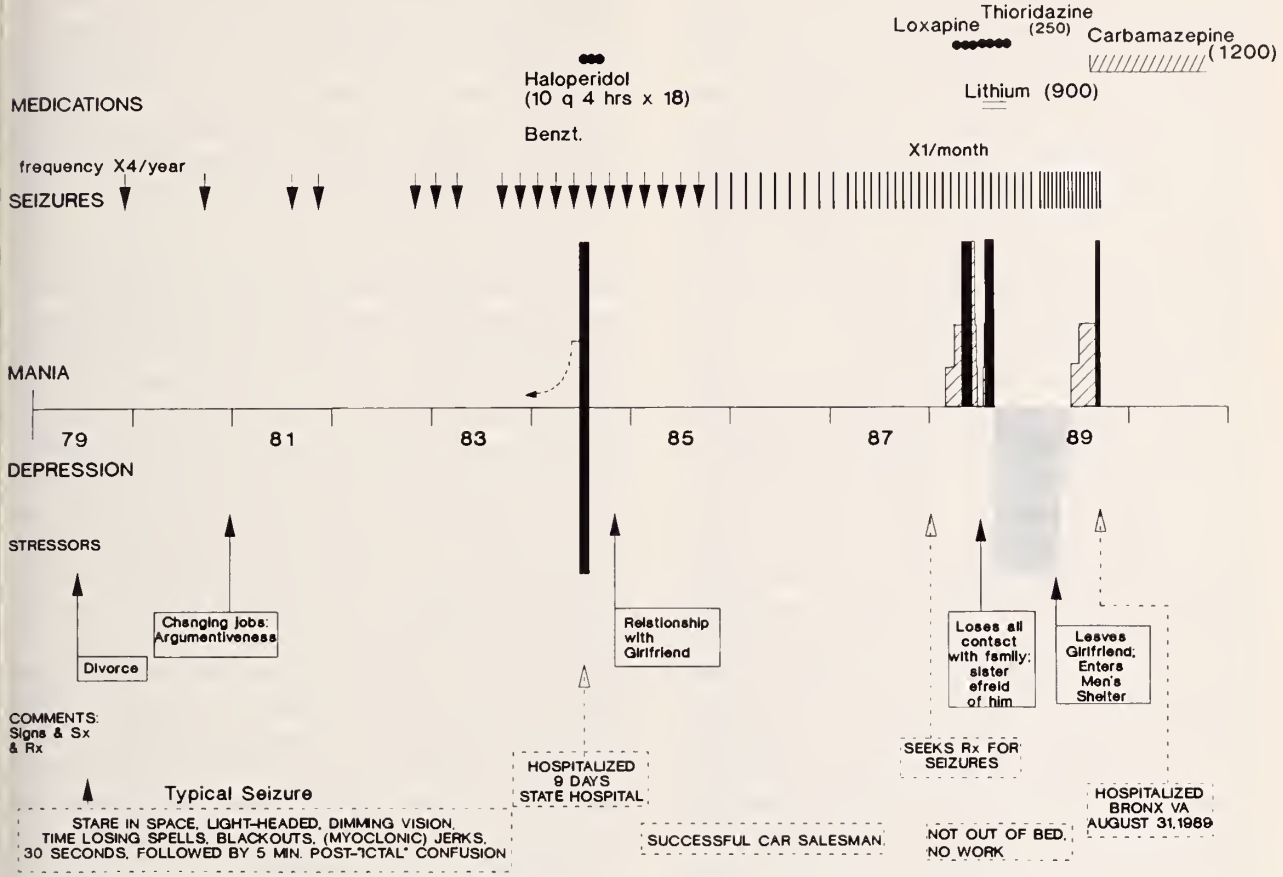
**Post:** Let's hear more about the seizures here.

## Hospital Course

**Findling:** Due to the patient's complaints of these time-losing spells, the patient had both an EEG and a sleep-deprived EEG with low temporal leads. The waking EEG showed anterior to mid-temporal sharp waves, more frequent on the right than on the left. The sleep-deprived EEG revealed right anterior to mid-temporal changes with some sharp left temporal changes. Although magnetic resonance imaging was requested, the patient felt "claustrophobic" during the test and was unable to complete it. The CT scan with contrast was normal.

During the time the patient was being evaluated for these "time-losing" spells, a seizure episode was noted. The patient was sitting in a community meeting when he developed clonic jerking of all four extremities for approximately 5 seconds. He was then disoriented for about 30 seconds, and felt well thereafter. He stated he had no recollection of what occurred during these episodes, but the subsequent disorientation he described as a typical spell. The patient was subse-

RELATIONSHIP BETWEEN SEIZURES AND BIPOLAR AFFECTIVE DISORDER: LIFECHART CLARIFIES SEQUENTIAL AND PROGRESSIVE EVOLUTION



quently started on carbamazepine. He had no further seizure episodes noted, with a subsequent stabilization in mood. The patient was maintained on 1200 mg/day of carbamazepine with a blood level between 8 and 12 µg/mL.

The patient was discharged to a men's shelter and was followed by an outpatient psychiatrist. Upon discharge, he was euthymic and intended to find a job as a car salesperson, again with Chrysler. During the four weeks after his carbamazepine dose was stabilized, the patient reported no more time-losing spells.

## Differential Diagnosis

**Post:** So he doesn't have spikes and waves. He's got sharp waves on his EEG. These sharp waves, were they spikes or just sharp waves? Because if they're anterior sharp waves, they don't mean anything. If they're clear-cut runs of spikes, I think people would feel more comfortable with a diagnosis of complex partial seizures; if they're clear-cut spike and wave patterns, one would be more willing to entertain absence-related episodes.

**Clinical Diagnosis: Complex Partial Seizures?** So, in fact, we have a problem with the EEG confirmation of complex partial seizures, but before we deal with that, we should acknowledge that the history is classical for complex partial seizures, and I think some neurologists would be willing to make the diagnosis on the basis of this clinical history. He's got staring, light-headedness, blackouts which last for 30 seconds, sometimes with jerking motions, and he's confused for five or ten minutes thereafter. These have been multiply observed and witnessed and—what is most interesting—they show a potential from initial occurrences of four times a year, to acceleration in frequency over ten years, to occurrence a couple of times a day. It would be nice if the EEG actually showed that he had classical complex partial seizures with spikes in his temporal lobe. The way I hear the EEG report is that it is not confirmatory, and, in addition, there are some problems with this onset.

Psychomotor epilepsy in a 40-year-old is not the typical time of onset. It's just like you're all led to the conclusion that this patient is not schizophrenic because he's had a good premorbid adjustment with a very competent set of relationships, and schizophrenia usually doesn't start at age 40. The same thing is true for idiopathic complex partial seizures—they are somewhat atypical for a 40-year-old onset. It makes one think of potential organic causes, including the possibility of a slow-growing tumor in his temporal lobe, which is just getting big enough to set up an irritable focus and trigger these spells, which are, in fact, accelerating dramatically.

**Problem of EEG Confirmation.** So our working diagnosis is that he's likely to have complex partial seizures based on clinical history, and yet we don't have that documented on the EEG. The question is, how much is that going to bother us, either as psychiatrists or as the treating and consulting neurologist—what is the consultant going to do about it? Our neurology consultant [in the audience] says that he's not going to worry about the EEG, and I would have to agree with him completely. But I can also tell you that I had a case presented to me two weeks ago at the National Institute of Mental Health in Consultation Rounds at which Dr. Roger Porter said that the patient didn't have a complex partial seizure disorder because the EEG was suspicious but was never verified. The spells were labeled pseudoseizures or hysterical seizures.

I would reverse the process in this current case and suggest that our patient has complex partial seizures until proven otherwise, because we know that any number of EEGs with provocative tests (sleep deprivation, hyperventilation) or with nasopharyngeal leads may not produce those spikes until a depth or subdural insertion of electrodes is done. In a nice case by Devinsky in the *Journal of Clinical Psychiatry* just this past year, an adolescent girl had what were "obviously" pseudoseizures because they were so hysteriform and involved with all sorts of bizarre actions; she had about 30 EEGs (6). Nothing ever showed up. They then inserted an electrode strip underneath the temporal lobe, and again nothing happened until the medial parts of the temporal lobe were reached, and then, all of a sudden, a spike focus came out during these "hysterical" seizures. So we have to be careful about using this EEG either to pin on somebody a nondiagnosis, or a diagnosis of hysterical seizures, without the appropriate level of diagnostic certainty and/or invasive studies.

**Status of Temporal Lobe.** I would also ask our neurologist now, is he worried that something is growing in this person's temporal lobe? In spite of the excellent response to carbamazepine (no more seizures), the etiology actually does need to be investigated further,. because the seizure process has been progressive, even if the etiologic lesion is not (that is, one could have a constant cyst, an arteriovenous malformation, or other static lesion still propelling an evolving ictal process).

Neither lumbar puncture nor magnetic resonance imaging has been done, because like many people with or without a psychiatric diagnosis, the patient had a lot of trouble sticking his head in this "atom smasher," or whatever it seemed like to him. So we don't know if he's got a tooth growing in his temporal lobe, if he's got a hamartoma there, if he's got a little vascular malformation, if he's bled in there. We're comforted, but only partially, by the fact that the putative lesion has been there for a long time without choosing other neurological signs. Yet it certainly has had a "malignant" clinical progression during this phase; if it's an irritable focus, it was getting a lot more irritable. Even though the spells were stopped with drugs, the question is, what's going to happen to this patient in the future? I would submit that there's some reasonable chance that he's going to start brewing up seizures through this carbamazepine treatment.

**Tolerance to Psychotropic Agents.** Neurologists don't like to talk about tolerance to the effects of anticonvulsant agents, but it happens, and tolerance also develops to the effects of psychotropic agents, which is not widely recognized. There's a "conspiracy" not to mention that, but even with drugs like lithium carbonate that work in manic-depressive illness and are supposed to work forever, you as clinicians know that that's not true. Don't you? You do! We all denied it for ten or fifteen years. Schou put the patients on lithium and they got all better. He followed them for a fairly long time, but formally in his systematic studies only for several years. In the most recent study by Maj et al. (7) of patients who were taking lithium and had complete remission for two years, almost half started having episodes, some with many recurrences, if they were followed for five years longer.

So even if one picks the complete lithium responders to start out with, as Maj did, some people will break through lithium treatment. This represents either progression of the underlying illness in spite of the pharmacology, or some sort of tolerance phenomenon. Later I'll suggest an interesting preclinical model, "contingent tolerance," that may be pertinent to the breakthrough of seizure episodes that I have postulated will recur. It is also possible that affective episodes may reemerge. Contingent tolerance may be pertinent to both types of breakthrough (8–10).

**Acceleration of Affective Episodes.** So we've got some notion about these spells and the problem of their differential diagnosis. At the same time, we've got clearcut affective episodes also brewing up in this patient. In 1984 he has his first one, when he's about 44 years old, and then he goes for four years without another one. Then he starts to have them repeatedly. On admission, he is continuously depressed—I think that was nine months he couldn't get out of bed. So he's actually having an acceleration in the evolution of his affective episodes as well. This doesn't happen in everyone, but Kraepelin observed it earlier in the century, and many others have. Patients who have a first affective episode are going to have others—80% of the time, even after first depressions (11). Even in the unipolar variety, Grof's data (12) suggest that the next episode is going to occur at some undetermined interval, but even if the interval is long (ten years between the first episode of depression and the next one) the next episode will come at a shorter interval, the next at an even shorter one, and so on. There is a clear acceleration in the course of these episodes, apparently reflecting an increasing vulnerability.

If the first episode is manic, the illness is almost invariably recurrent (2, 11, 13), often in an accelerating pattern. I think that's something we have to be thinking about when we're treating a patient acutely, either with a first or second episode; we should be asking, do they need to be on prophylaxis to prevent what otherwise can be a malignant deterioration of affective illness? Let's just imagine a patient without any seizures complicating the affective picture for the moment. The issue is whether we should start treating our affectively ill patients prophylactically earlier to prevent recurrences and the potential increased vulnerability that accompanies them.

I would submit that in the person who has a first early manic episode severe enough for hospitalization and has a family history of affective illness, you ought to think about prophylaxis different from what most textbooks advise. I don't know how many people would initiate prophylaxis following a first severe manic episode in a patient with a loaded family history. Does anybody think about that with such a patient and his family? Yes, I would think that that's worth thinking about. (See Strober et al [14] for the high risk of relapses in 18 months after a light episode.) That's different from a first minor depressive episode in a young person, who might have ten or twelve years before another episode; one wouldn't want to start prophylaxis at that point.

So if the first or second episode was potentially catastrophic in its consequences, if the patient was acutely suicidal, again with a loaded family history, you might want to consider prophylaxis. Based on the recent data of Montgomery (15) and Rouillon et al. (16), there is a high risk of relapse in the year following the discontinuation of acute antidepressant treatment of patients with even a few prior episodes.

**Progression from Psychosocial Triggering to Autonomy.** Another interesting issue that we have to deal with here is that before any of these episodes started, including the spells, there were many potential psychosocial precipitants. The patient lost a spouse, and then we know he feels terrible about the loss of his family, and he's also lost his girlfriend. He's undergone a social deterioration. One can say, well, it's just because he's getting these spells—he's got complex partial seizures and his brain's firing off with an autonomous spike focus, and it's deranged his social relationships. But the social stressor came first.

The data now are very convincing that, given a look at the first episode of affective illness, be it manic or depressed, and a good history, life events are implicated in a high proportion of those first episodes (17). There's a statistically significant lesser likelihood of a prominent psychosocial precipitant for the later or recurrent episodes in both unipolar and bipolar illness. There are now twelve such studies showing a higher incidence of psychosocial precipitants with the first compared to subsequent episodes. The onset of complex partial seizures can also be triggered by stress. Our neurological consultant seems to be less worried about psychosocial precipitants of epilepsy than we are of affective illness.

There's no question that stresses can be associated with the onset or recurrence of seizures. But some of our biological colleagues think affective illness only emerges out of the blue. Other clinicians think the process can only be psychological; the field has dichotomized to one or the other. I would submit the choice doesn't have to be made; the options are not mutually exclusive. In many patients, negative things are going on in their lives at the beginning of the illness that precipitate depressions or manias and then, as episodes recur, the illness becomes more autonomous. The biochemical pathways become facilitated to the point that episodes are occurring without psychosocial triggers.

This conceptualization puts together a psychodynamic context for episodes that may be triggerable in the beginning and a biological-endogenous context for episodes that become spontaneous and automatic at the end (17–19). We know from a kindling model that that's exactly what happens in a literal model of kindling; that is, if you take an animal and you put an electrode down into the amygdala of the temporal lobe, a spike focus eventually develops. You stimulate the amygdala once a day for one second in the rat and nothing happens initially. You keep on stimulating the amygdala once a day for a while and then, all of a sudden, you start to exceed the afterdischarge threshold and you start to see runs of spikes in the amygdala. The spikes or afterdischarges get longer, more complex, and widespread, and then, finally, the animal begins to develop full-blown stage 4 seizures in response to the same one-second stimulation that didn't produce any effect in the beginning. This is the development phase of kindling.

This phase involves two things—a lowering of the afterdischarge threshold such that the animal is more vulnerable to having these spike runs and motor seizure evolution. The animal progresses from stimulus-induced behavioral arrest and staring episodes (stage 1 & 2), to unilateral forepaw involvement (stage 3), to bilateral forepaw involvement (stage 4), and rearing and falling in a full-blown, stage 5 seizure. So you have kindled a threshold excitability change and you've made the animal have seizures to the same once-daily one-second stimulations that produced nothing in the beginning. That's kindling development.

Continue this repeatedly, about fifty times, and every time you stimulate the amygdala, the animal has a seizure. Then, one time you observe that the animal in the cage is having a seizure without any electrical stimulation. So you've kindled this animal from (A) the development phase to (B) the completed phase to (C) the spontaneous phase. This is obviously a good model for spontaneous epilepsy.

I would submit that even in a classical bipolar patient with no seizure focus, kindling evolution to spontaneity is an interesting analogy for the evolution of affective episodes, from the phase in which there is early psychosocial stress that doesn't produce any affective disorder, to the phase in which repeated separations, losses, or other stressors in that vulnerable individual are going to be associated with the development of full-blown affective episodes. If enough affective episodes that are precipitated occur, vulnerability may increase to the point where episodes begin to occur on their own. If seizures can migrate from precipitated to autonomous, why cannot the same sort of progression occur in affective disorders; i.e., from psychosocially triggered to autonomous?

**Analogous Evolution in Response to Drugs.** In this kindled seizure model, not only does the process evolve over time, but something else is startling. The effectiveness of pharmacological intervention varies as a function of stage of evolution (17). That is, treatment response in the early development phase, the middle completed phase, and the late spontaneous phase is totally different. For example, carbamazepine (Tegretol), the drug that our patient happens to be responding to for both his affective illness and his "spells," will not prevent the early development of amygdala kindling in the rat. Yet it works just fine on the middle, completed phase (we don't know about the late spontaneous phase). Another anticonvulsant, diazepam (Valium) works on development and completed phases, but not on the spontaneous phase. Conversely, phenytoin (Dilantin) shows the opposite pattern; it doesn't work well early, but as seizures become spontaneous, it's effective (20, 21).

Different drugs are picking off different stages in the evolution of this kindling process, implying a change in the neural substrates and/or biochemistry of the process. When the seizures have evolved to the spontaneous form, why shouldn't parenteral diazepam work anymore, since it's a potent central-type benzodiazepine receptor agonist? Diazepam blocks kindling development and completed seizures, but it won't block the spontaneous ones, suggesting that the latter type of seizures must be differently encoded. There's a different biochemistry at the end of this process than at the beginning, and it's not benzodiazepine-responsive any more. So the biochemistry and/or the anatomy has migrated to something that's not benzodiazepine-responsive; all of a sudden it picks up phenytoin responsivity.

I would wonder with you whether an analogous evolution in pharmaco-responsivity doesn't also occur in the course of affective illness. And again, we put this patient aside for the moment, since he complicates things because he's got both seizures and affective disorders. But in classical primary affective illness, initial minor depressions and stress-related dysphorias may have a differential responsivity from the later, major ones (17). I think we'd all pretty much agree that there is a differential responsivity, and that the minor ones respond to psychotherapy very nicely. The NIMH Collaborative Study redocuments what we all knew—that the minor depressions do as well with psychotherapy as they do with typical antidepressant drugs, but severe major depressions and psychotic depression do not respond well to psychotherapy alone. For severe major depressions, tricyclic antidepressants and monoamine oxidase inhibitors are needed. For delusional patients, neuroleptics or electroconvulsive therapy may also be needed. Treatment later in the course of illness may need to differ from earlier treatment, when psychotherapy or even low-dose benzodiazepines may be effective. Nobody likes to talk about benzodiazepines because they're abused and overprescribed for affective illness, but the early studies by Rickels and others (22) indicated that minor outpatient depressions were also benzodiazepine-responsive. The more severe major ones may not be responsive to benzodiazepines.

Lithium is also highly effective in the mid-phase of illness, but a large series of studies show that by the time rapid cycling appears (usually late in the course of illness), lithium is less effective (23). Instead of an 80% response rate to lithium, as in the early or middle phase, the response rate is much lower in the late rapid cycling, spontaneous variety. Some of the anticonvulsants such as carbamazepine or valproate are effective in the late phases. Some patients respond to carbamazepine and not valproate, and vice versa (24). So one negative trial with an anticonvulsant cannot predict refractoriness to another. The clinician has to systematically assess whether the patient is going to respond to carbamazepine. Moreover, although the tricyclics and the MAOIs are excellent for severe middle phase depressions, those agents can trigger manias or accelerate cycle frequency in some bipolars and rapid cyclers, as Wehr, Goodwin, Kukopulos, and many other people have indicated (25).

So there may be some change in psychopharmacology as a function of stage of evolution of affective illness, just as there is with kindling. While our patients with primary affective illness don't seize, we can still use this indirect kindled seizure analogy to help us focus on the question of possible differences in pharmacology as a function of illness evolution.

**Longitudinal Changes in Psychotherapies and Pharmacotherapies.** If the spontaneous seizure has migrated somewhere, where has it gone? An investigator named Pinel in Canada kindled animals to spontaneity, then looked at their seizure focus, which had initially been in the cortex, at the tip of the stimulating electrode in this case (20). During spontaneous seizures the focus wasn't there any more. It had disappeared—another reason for not accepting a negative EEG as an indication that the spell is not a true seizure disorder. During spontaneity, rats aren't having hysterical seizures. The seizures are real, but the focus is either suppressed or (the more startling possibility) has gone somewhere else—it has migrated. We don't know where it is, and I would submit that analogous changes could also occur in affective illness. With each successive episode, as the liabilities for further depressions and manias increase and the biological changes get more and more ingrained, the neural pathways and programs for them may be evolving and migrating as well.

We should think about this in terms of our psychotherapies as well as our pharmacotherapies. Dynamic psychotherapy around the early stresses and stress-related dysphorias may be the appropriate thing. But if one has enough of these affective episodes and they're happening relatively automatically, maybe we should actually be shifting to cognitive and behavioral therapy modes, using systematic desensitization and the like to get more at the automatic triggers that make the episode look spontaneous. Mort Mishkin has some beautiful data that show how renewal substances of memory change in part as a function of repetition of the process (26). Take out an animal's amygdala and hippocampus, and the animal's ability to use "representational" memory is gone. But train the animal to do a task over and over again, and the task gets into the "habit memory system" (as Mishkin calls it), which is not impaired by amygdala and hippocampal lesions any more. The memory trace is now no longer dependent on the limbic system and has apparently gone into the striatum (the tail of the caudate); it has now become an automatic process.

Mishkin's studies imply that the anatomic substrates for memory in primates are changing as a function of repetition. The famous patient H.M. also demonstrates that this happens in humans. H.M. has no medial temporal lobe structures on either side, and he can't remember the doctor whom he has just seen or what he has eaten for breakfast that day. He has no representational memory. But, in fact, he knows how to do the tests that the doctor has him do every day. He doesn't know he knows, but he just goes ahead and does the tests, and he does them correctly because they are in his habit system. Perhaps this substrate has changed over time; clearly it is not limbic-dependent.

I would submit that as affective episodes are repeated enough, they may be getting more and more ingrained and automatic in our habit system. We have to think of ways of trying to stop the progression of the illness so that it does not become automatic. This should include earlier institution of pharmacoprophylaxis and its maintenance for an indefinite period of time, perhaps for the patient's whole life. When we see patients who are in later phases of the course of their illness, perhaps we need to think about revising our therapies, both pharmacotherapies and psychotherapies, so they are appropriate to where the patients are in the evolution of their illness process. Dynamic therapy targeting representational-type mechanisms may no longer be appropriate, and we may have to develop more behavioral approaches to target habit-type memory mechanisms.

**Life Stresses, Stress Hormones, and Sensitization.** One other concept may be pertinent to our patient. Divorce and related stressors may have played a role in the initial processes of this patient's seizure and affective disorder. Some mechanisms available to us from preclinical science may help explain stress sensitization to the point where changes become automatic. Putting stress hormones into the amygdala (once a day, every day) can actually kindle seizures to stress hormones. That's been done with the opiates by an investigator named Cain in Canada (27). Put some opiates in the amygdala every day, and the animals will kindle seizures, not in the beginning, but eventually; then the seizures can, with enough repetitions at the end, become spontaneous. Put some corticotropin-releasing hormone (CRH) into the cerebral ventricles of experimental animals, and they too will have late-onset seizures (28, 29).

So some human stress hormones could, hypothetically, be playing a role in either the seizure sensitization or the affective episode sensitization in our particular patient. Affective illness is associated with cortisol hypersecretion in at least 50% of patients, and some of that may be driven by changes in stress hormones. An excess CRH in cerebrospinal fluid has been found in some, but not all, studies of depressed patients (30). However, there is other evidence that even when CRH is not elevated in the cerebrospinal fluid of depressed patients, evidence of cortisol hypersecretion—from dexamethasone suppression test or a blunted ACTH response to CRH (31)—is still available, indicating some sort of a change in the set point, such as the chronic hyperfunction of the hypothalamic-pituitary-adrenal axis that seems to occur in many patients along with their affective disorders. We can think about the possibility that some of the endocrine and peptide changes could be contributing to the behavioral and affective sensitization. The possibility is all the more enhanced by the recent finding in our laboratory that not only stressors (32, 33) but also CRH induces the proto-oncogene c-*fos* (34), which may provide a mechanism by which the environment can have an impact on gene expression (18).

So both psychoendocrinology and molecular biology are relevant to this environmental-neurobiological interface. We can thus think about the possibility that our patient's divorce either had a triggering effect or made the patient more vulnerable to seizures—seizures that were going to occur anyway because he has this small, slow-growing meningioma or other insult in his temporal lobe. It is also possible that he has a stress peptide-related seizure disorder that may not show up with classical spikes on the EEG and may produce atypical seizures.

**Contingent Tolerance.** The concept of contingent tolerance may be pertinent to the potential for breakthrough episodes. Work on this concept has been pioneered in our laboratory by Susan Weiss (8, 9). If the animal is kindled up to the phase when it is reliably having complete seizures, and then given carbamazepine prior to the stimulation, it doesn't have a seizure. This is expected, because carbamazepine is effective on completed seizures. If carbamazepine is given again the next day, it is still effective. However, if the animal is kindled and given carbamazepine before stimulation for five or six days, the animal will become tolerant to carbamazepine and start having seizures. Carbamazepine is still being administered every day, and the animal is seizing right through it. So the researcher says, Oh, well it's just tolerance, something's wrong with the blood levels. But that's not the case at all, because if a seizure is produced in the animal and, after the seizure each day, the animal is then given carbamazepine, the animal does not become tolerant. So the tolerance is dependent on giving the animal the drug while it's experiencing the seizure—on the pairing of the seizure and the drug (8, 9). If the drug and the seizure are unpaired—that is, the animal has a seizure and the drug is administered afterwards—tolerance does not occur.

Thus the tolerance is not pharmacokinetic. Further proof of that is that once these animals become tolerant (having full-blown stage 5 seizures) and experience five days of seizures when the drug is given after the seizure, they will rerespond if the drug is then given prior to the kindling stimulation. Or, if the drug is withdrawn—that is, seizures are produced but no drug is administered—they rerespond. That's why we call the phenomenon contingent tolerance—it is dependent on the contingencies of the pairing of the drug with the episode that is being treated.

In a parallel fashion, I surmise that if and when our patient begins to have his complex partial seizures or his affective episodes brewing up through his successful anticonvulsant treatment (somebody ought to be following him to make sure he's not beginning to have seizures again) and he is taken off the drug for a while, analogous to the preclinical model, he might also rerespond.

Once rats are tolerant to carbamazepine, they can also be switched to another anticonvulsant that doesn't show cross-tolerance, such as diazepam (9). Animals will respond to benzodiazepines at this point, even though they have become tolerant to carbamazepine, but they will also eventually become tolerant to benzodiazepines, so one might use and rotate drugs that don't show cross-tolerance.

There is now an increasing literature about tolerance not only to carbamazepine in manic-depressive patients (35), but even to the antidepressant effects of the monoamine oxidase inhibitors (MAOIs). Baldessarini (36) has reported it, Donaldson has reported it (37), a number of others have written case reports. Somebody is doing well on MAOIs for a nice long time, then starts having depressions through the drug. Jan Fawcett and a number of other people say that if MAOI treatment is stopped for a while and the patient has some depression off medication, resumption of MAOI treatment may re-institute the response.

If these observations are confirmed, contingent tolerance may be relevant not only for seizure disorders, but also for the management of loss of responsivity in the affective disorders.

### The Longitudinal Approach

My parting admonition is that you think about the longitudinal course of any patient you are treating. If you can, work out a visual map (see Fig. and ref. 1–3). It is really quite easy to do once you develop a systematic approach, and it allows you to line up the drugs on the top, the episodes in the middle, and the psychosocial stressors on the bottom. It helps organize the course of illness both for you and the patient, and it has many positive spinoffs in the end, particularly for the manic patient who has had three or four of these episodes and doesn't think that there's anything wrong with him. If you start to map these and she or he sees that they're occurring at whatever frequency they are, it will help a lot with compliance—bringing the patient into the therapeutic alliance.

Think about the course of the illness, and think about it in terms of both your psychotherapies and your pharmacotherapies, and with the idea that things don't stay the same over time—that your perspective, your snapshot of the patient who comes to see you one day in your office, may or may not reflect what will happen to this patient over the next several years. Refocus on this long-term perspective and what you have to do to maximize the chances that you're really going to have the illness under control in the long run.

I have one additional cautionary note. At NIMH we tend to see the failures, the refractory patients, and so perhaps my advice is more conservative than others. But I would say that if somebody is doing well on a given regimen, lithium or whatever, you really ought to leave that regimen in place, or perhaps risk the grave danger of the patient's having recurrences. Conversely, if the patient is not doing well, be aggressive, change things around.

Another reason for conservatism in maintaining prophylaxis is based on a newly described phenomenon—one we would not have uncovered if we were not life-charting our patients. We have seen patients who have been stable on lithium for 6, 7, even 15 years, then stop taking their lithium, with or without their physician's support and acquiescence. They then have a severe episode, restart their lithium, but never respond again to it, or even to lithium plus adjunctive treatments (38). We've seen that now in four of our patients. I've talked to Jules Angst, who has hundreds of patients in his clinic in Switzerland, and he says that not only do more than 80% of his lithium-responsive patients relapse when they stop taking lithium, but a substantial portion do not rerespond as well again. This absence of reresponse is an additional reason for conservatism in the longitudinal approach to recurrent affective disorder. If you've got the illness under control, give strong consideration to leaving well enough alone. Think twice before removing the therapeutic agent. If you don't have the disorder under control—if the patient is having episodes in spite of what you're doing—be radical and think about revisions in the therapy.

We now have many different antimanic treatments (24), whereas two decades ago we had only lithium and neuroleptics. With the appropriate mixing and matching of treatments, I think we can do much, even for the most treatment-refractory bipolar patient. When the patient has two, perhaps interrelated, processes—seizures and affective illness—the diagnostic and therapeutic complexities and greatly increased (39, 40). But the potentials for successful management (even with a single agent) are there, as highlighted by this case (see Fig.).

## References

1. Squillace K, Post RM, Savard R, Erwin M. Life charting of the longitudinal course of recurrent affective illness. In: Post RM, Ballenger JC, eds. Neurobiology of mood disorders. Baltimore: Williams & Wilkins, 1984:38–59.
2. Roy-Byrne PP, Post RM, Uhde TW, Porcu T, Davis DD. The longitudinal course of recurrent affective illness: life chart data from research patients at NIMH. Acta Psychiatr Scand (Suppl) 1985; 71:5–34.
3. Post RM, Roy-Byrne PP, Uhde TW. Graphic representation of the life course of illness in patients with affective disorder. Am J Psychiatry 1988; 145:844–848.
4. Carpenter WT Jr, Strauss JS, Muleh S. Are there pathognomonic symptoms in schizophrenia? An empiric investigation of Schneider's first-rank symptoms. Arch Gen Psychiatry 1973; 28:847–852.
5. Abrams R, Taylor MA. Importance of schizophrenic symptoms in the diagnosis of mania. Am J Psychiatry 1981; 138:658–661.
6. Devinsky O, Sato S, Theodore WH, Porter RJ. Fear episodes due to limbic seizures with normal ictal scalp EEG: a subdural electrographic study. J Clin Psychiatry 1989; 50:28–30.
7. Maj M, Pirozzi R, Kemali D. Long-term outcome of lithium prophylaxis in patients initially classified as complete responders. Psychopharmacology (Berlin) 1989; 98:535–538.
8. Weiss SRB, Post RM. Development and reversal of conditioned inefficacy and tolerance to the anticonvulsant effects of carbamazepine. Epilepsia 1991; 32:140–143.
9. Weiss SRB, Post RM. Contingent tolerance to carbamazepine: a peripheral-type benzodiazepine mechanism. Eur J Pharmacol 1991.
10. Pazzaglia PJ, Post RM. Contingent tolerance and re-response to carbamazepine: a case study in a patient with trigeminal neuralgia and bipolar disorder. J Neuropsychiatry Clin Neurosci 1991 (in press).
11. Zis AP, Goodwin FK. Major affective disorder as a recurrent illness. Arch Gen Psychiatry 1979; 36:835–839.
12. Grof P, Angst J, Haines T. The clinical course of depression: practical issues. In: Schattauer FK, ed. Symposia medica hoest: classification and prediction of outcome of depression, vol. 8. New York: Schattauer, 1974:141–148.
13. Goodwin FK, Jamison KR. Manic-depressive illness. New York: Oxford University Press, 1990:
14. Strober MT, Morrell W, Lampert C, Burroughs J. Relapse following discontinuation of lithium maintenance therapy in adolescents with bipolar I illness: a naturalistic study. Am J Psychiatry 1990; 147:457–461.
15. Montgomery SA, Dufour H, Brion S, et al. The prophylactic efficacy of fluoxetine in unipolar depression. Br J Psychiatry 1988; 153:69–76.
16. Rouillon F, Phillips R, Serrurier D, Ansart E, Gerard MJ. [Recurrence of unipolar depression and efficacy of maprotiline]. Encephale 1989; 15:527–534.
17. Post RM, Rubinow DR, Ballenger JC. Conditioning and sensitization in the longitudinal course of affective illness. Br J Psychiatry 1986; 149:191–201.
18. Post RM. The transduction of psychosocial stress into the neurobiology of recurrent affective disorder. Am J Psychiatry 1991.
19. Post RM, Rubinow DR, Ballenger JC. Conditioning, sensitization, and kindling: implications for the course of affective illness. In: Post RM, Ballenger JC, eds. Neurobiology of mood disorders. Baltimore: Williams & Wilkins, 1984:432–466.
20. Pinel JPJ. Kindling-induced experimental epilepsy in rats: cortical stimulation. Exp Neurol 1981; 72:559–569.
21. Pinel JPJ. Effects of diazepam and diphenylhydantoin on elicited and spontaneous seizures in kindled rats: a double dissociation. Pharmacol Biochem Behav 1983; 18:61–63.
22. Rickels K. Antineurotic agents: Specific and non-specific effects. In: Efron DH, Cole JO, Levine J, Wittenborn JR, eds. Psychopharmacology: a review of progress 1957–1967. Washington DC: US Government Printing Office, 1968:231–247.
23. Post RM, Kramlinger KG, Altshuler LL, Ketter T, Denicoff K. Treatment of rapid cycling bipolar illness. Psychopharmacol Bull 1990; 26:37–47.
24. Post RM. Alternatives to lithium for bipolar affective illness. In: Tasman A, Goldfinger SM, Kaufmann CA, eds. Review of psychiatry, vol. 9. Washington DC: American Psychiatric Press, 1990:170–202.
25. Wehr TA, Goodwin FK. Rapid cycling in manic-depressives induced by tricyclic antidepressants. Arch Gen Psychiatry 1979; 36:555–559.
26. Mishkin M, Appenzeller T. The anatomy of memory. Sci Am 1987; 256:80–89.
27. Cain DP, Corcoran ME. Intracerebral beta-endorphin, met-enkephalin and morphine: kindling of seizures and handling-induced potentiation of epileptiform effects. Life Sci 1984; 34:2535–2542.
28. Ehlers CL, Henriksen SJ, Wang M, Rivier J, Vale W, Bloom FE. Corticotropin releasing factor produces increases in brain excitability and convulsive seizures in rats. Brain Res 1983; 278:332–336.
29. Weiss SR, Post RM, Gold PW, Chrousos G, Sullivan TL, Walker D, Pert A. CRF-induced seizures and behavior: interaction with amygdala kindling. Brain Res 1986; 372:345–351.
30. Nemeroff CB, Widerlov E, Bissette G. Elevated concentrations of CSF corticotropin-releasing factor-like immunoreactivity in depressed patients. Science 1984; 226:1342–1344.
31. Gold PW, Goodwin FK, Chrousos GP. Clinical and bio-

chemical manifestations of depression: relation to the neurobiology of stress. N Engl J Med 1988; 319:348–353.
32. Daval JL, Nakajima T, Gleiter CH, Post RM, Marangos PJ. Mouse brain c-fos mRNA distribution following a single electroconvulsive shock. J Neurochem 1989; 52:1954–1957.
33. Nakajima T, Post RM, Weiss SRB, Pert A, Ketter T. Perspectives on the mechanism of action of electroconvulsive therapy: anticonvulsant, dopaminergic, and c-fos oncogene effects. Conv Ther 1989; 5:274–295.
34. Clark M, Weiss SRB, Post RM. Expression of c-fos mRNA in rat brain after intracerebroventricular administration of corticotropin-releasing hormone. Neurosci Lett 1991 (in press).
35. Post RM, Leverich G, Rosoff AS, Altshuler LL. Carbamazepine prophylaxis in refractory affective disorders: a focus on long-term followup. J Clin Psychopharmacol 1990; 10:318–327.
36. Baldessarini RJ. Current status of antidepressants: clinical pharmacology and therapy. J Clin Psychiatry 1989; 50:117–126.
37. Donaldson SR. Tolerance to phenelzine and subsequent refractory depression: three cases. J Clin Psychiatry 1989; 50:33–35.
38. Post RM, Leverich G, Altshuler L, Mikalauskas K. Lithium discontinuation-induced refractoriness. 1991 (unpublished manuscript)
39. Post RM, Uhde TW, Joffe RT, Bierer L. Psychiatric manifestations and implications of seizure disorders. In: Extein I, Gold M, eds. Medical mimics of psychiatric disorders. Washington, D.C.: American Psychiatric Association Press, 1986:35–91.
40. Post RM, Altshuler LL, Ketter T, Denicoff K, Weiss SRB. Antiepileptic drugs in affective illness: clinical and theoretical implications. In: Smith DB, Treiman DM, Trimble M, eds. Advances in neurology, vol. 55. New York: Raven Press, 1991:239–277.

# *Grand Rounds In Psychiatry*

# Acquired Brain Lesions and Psychiatric Illness:

## Possible Role of Dopaminergic Systems

Thomas A. M. Kramer, M.D., Arnold Merriam, M.D., Philip Harvey, Ph.D., and Michelle D. Lisoskie, M.D.

### Case Report

A 38-year-old, divorced, Jamaican, Afro-American man, a former chef and caterer who was unemployed and was living with his sister, came to the hospital in the fall of 1989 saying "I have no blood or water in my body."

Before July 1985 he ran a Chinese restaurant, catered social events in his spare time, and lived with his grandmother. He is described by family members as a man who was self-assured, strong-willed, a person who "loved the world," and who enjoyed relationships with women. The patient and family members reported no previous alcohol or drug abuse. Until July 1985, he displayed no signs of psychiatric illness.

In July 1985, he was stabbed through the left parietal skull with an icepick wielded by a thief who was burglarizing his grandmother's home. Neurosurgery was delayed owing to his grandmother's reluctance to sign consent for the procedure.

The patient recovered from the procedure and was evaluated two weeks later by a speech and language pathologist, who found him to be "in a heightened state of activity with severely decreased ability to process information." Hearing was "adequate for communication purposes." Visual acuity of the right eye was noted to be impaired. The patient became hostile and discontinued rehabilitation after his third visit.

The patient's family corroborated these findings. The family had the patient tutored at home, and he regained the ability to recognize letters and numbers and eventually to read. He was prescribed phenobarbital after his hospitalization but he stopped taking his medication.

During the months following his injury, the family also noted a change in the patient's personality. He became threatening and abusive toward his grandmother. He armed himself with knives and machetes. Plates, silverware, and kitchen appliances were stored away behind his padlocked bedroom door. The patient complained of "things coming out of the walls and the floor" and dug holes in them. Family and friends stopped visiting because the patient would provoke arguments with them and threaten them. He initiated a physical altercation with one neighbor. His sister, concerned about her grandmother's safety, arranged for him to move to an aunt's house. After staying there a short time, in the spring of 1986, the patient decided to visit friends in Canada.

He arrived at his sister's house in New York unexpectedly (a few days after he had embarked

---

Adapted from the authors' Grand Rounds in Psychiatry presentation at the Mount Sinai Medical Center.

Case Presenter: Michelle Lisoskie; Neuropsychological Assessment: Philip Harvey; Case Discussant: Arnold Merriam; Editor: Thomas A. M. Kramer.

From the Division of Residency Training and Education, Department of Psychiatry, Mount Sinai School of Medicine (CUNY), New York, NY (TAMK, PH, ML), and the Division of Neuropsychiatry, Department of Psychiatry, Albert Einstein School of Medicine (AM), Bronx, NY. Address reprint requests to Thomas A. M. Kramer, M.D., Director of Psychiatric Outpatient Services, University of Arkansas for Medical Sciences, Mail Slot #568, 4301 W. Markham, ARK 72205.

for Canada) and moved into the lower floor of her house. His behavior became more bizarre; he grew a long beard and refused to cut his hair. Around his bed he kept pictures of relatives who had passed away and began accusing family members of killing them. He continued to get into many arguments. For hours he stood by a window and read the Bible; he sent Christmas cards signed "Jesus." He was suspicious of food cooked for him by family members and verbally abusive to anyone who suggested he modify his behavior. The family was convinced that the patient needed psychiatric treatment.

After a court order remanded him for psychiatric care, in January 1987 he was admitted to Bronx Municipal Hospital Center.

At the time of discharge in April 1987, the patient was followed up at a psychiatric outpatient clinic and maintained on antipsychotic medication. He began neurologic rehabilitation at Mount Sinai Medical Center's Department of Rehabilitation Medicine, and through the Center he was enrolled in a cooking school for mentally impaired patients.

In September 1988, rehabilitation workers noted that the patient was depressed after a visit to his family and friends in Canada; he was referred to a psychiatrist. He admitted to depressed mood, decreased sleep and appetite for several months, and feeling suicidal without a plan. He complained of "blank spells" and increasing difficulty with his memory for a month prior to the consultation. He denied any psychotic symptoms. The family reported that the patient slept well but that he refused to believe that himself. Electroencephalogram with nasopharyngeal leads at the time showed no epileptogenic focus. Therapy with the antidepressant fluoxetine was begun, and family members agreed that his mood seemed to improve.

The patient graduated from cooking school in early 1989 and was placed in a job in a nursing-home kitchen. The patient's sister showed him a few times how to travel by public transportation to his job. However, he left the job after two months because he could not consistently recall which buses to take to get to work. After leaving the job, he continued to complain of difficulty sleeping, and his depressed mood worsened. He refused enrollment in a work rehabilitation program and spent the summer of 1989 at home, watching television. Fluoxetine was discontinued in July 1989, several months after recurrence of depressive symptoms. Antipsychotic medication was tapered because of dystonia and abnormal tongue movements consistent with tardive dyskinesia. However, in September 1989 the patient's sister called to inform his psychiatrist that the patient was experiencing delusions that he had "no pulse . . . my blood is dried up." He was referred to the Mount Sinai emergency room for possible admission secondary to psychotic symptoms.

**Neuropsychological Assessment.** A neuropsychologic assessment was undertaken to determine the nature and extent of the patient's dysfunction and to estimate his level of intellectual functioning. His full-scale Wechsler Adult Intelligence Scale (WAIS-R) IQ was 61, based on nearly identical verbal and performance IQs of 61 and 67, respectively. Since the patient scores in the mentally retarded range on these assessments, it is clear that his abilities are notably impaired relative to his estimated premorbid level, which was at least average.

A number of additional tests were performed to identify functional impairments, since the patient was reported clinically to manifest signs of aphasia and memory deficit. The evaluation was also intended to clarify whether deficits from the parietal/temporal area were localized to the side of the lesion. The patient was examined with the Benton-Indiana Aphasia Screening Inventory and found to have signs of receptive (fluent) aphasics that were pronounced, but there were more limited signs of nonfluent aphasics. Confrontation naming was notably impaired in comparison to the ability to generate responses to open-ended questions.

Memory span was very poor; auditory serial recall was correct on only half the trials of three digits, and the patient had no ability to recall even two digits backward. Copying of visual material was adequate except for notable neglect of the right visual field. The patient performed well above his IQ level on recall of visual stimuli, except for tendencies to neglect the right field.

Trailmaking ability was at the bottom of the brain-damaged range. Tactile recognition was totally absent for the right (contralateral to the lesion) hand, with no impairment on the left hand. Motor-echoing tasks were performed without error by the left hand and with no accuracy on the right hand.

Finally, the patient manifested very impaired performance on a test of prefrontal cortical functioning—the Wisconsin Card Sorting Test. The patient made numerous perseverative errors, although he completed one category. The completion of one category rules out a variety of comprehension and motivational variables and suggests that the perseverative tendencies were a function of a frontal functional deficit.

**Summary.** The patient manifested lateralized deficits in parietal, temporal, and, to a lesser extent, frontal brain functions. The more anterior the frontal task, the better it was performed. We interpret these data as suggesting that the neuropsychological dysfunction includes regional deficits in the area of the specific lesion as well as deficits associated with connective disruptions to other cortical areas. At the time of testing, the patient appeared generally aware of his impairments and was not disoriented. We conclude from this assessment that severe regional tissue loss has produced local deficits and impaired secondary associational functions, without inducing a generalized dementia.

## Discussion

Although the particular characteristics of this patient's prior life history, belief systems, personality, skills, ambitions, incapacities, frustrations, and coping strategies, and even the idiosyncratic details of his psychotic manifestations, are vital to his psychotherapeutic treatment and vocational guidance, these are not the aspects of the case that beg for explanation. Rather, the case requires asking important questions about the relationship between the patient's focal brain lesion and his psychotic illness.

Specifically, can an acquired brain lesion biologically induce a psychotic condition? If so, what is the means by which this occurs? What is the relationship between the brain region affected by the lesion and the psychiatric disorder? What is the relationship of psychiatric symptomatology to the neuropsychological dysfunction attendant on the lesion? What implications does the phenomenon of brain-lesion-induced psychosis hold for understanding basic brain-behavior relationships? What are the implications for the understanding of the idiopathic psychoses, and their nosology?

The literature suggests that brain injury is in fact a risk factor for the development of psychotic disorders. Davison and Bagley (1), in their review of schizophrenia-like states associated with central nervous system disease, concluded that brain damage was capable of producing psychotic illnesses that cannot reliably be distinguished from idiopathic schizophrenia. They furthermore contended that these disorders were true secondary psychoses in that they could not be explained on the basis of genetic or premorbid vulnerability and were associated in particular with lesions in the temporal lobe and diencephalon.

Achte et al. (2) evaluated 3,552 veterans who had sustained traumatic brain injury for the subsequent occurrence of psychotic disorders over a period of 22–26 years. They reported that 8.9% of their sample developed such illnesses, and were unable to detect a relationship between a schizophrenia-like illness and any particular lesion site. In over 76% of those who developed this complication, there was an interval of a year or more after the brain injury until the first psychotic break.

It is therefore apparent that brain injury is an independent risk factor for psychosis and that the appearance of psychiatric pathology is often delayed.

Anomalous behavior after brain injury is in some instances explicable on the basis of interference with fundamental neuropsychologic processes. Perhaps the best examples of this direct relationship between cognitive and behavioral dysfunction are the distinctive syndromes observed in the wake of prefrontal brain damage, where the defects in overt behavior can be reliably ascribed to such neuropsychological deficits as impairments in inhibition, initiation, and execution of sequential motor and cognitive operations (3).

Our patient's psychosis, however, cannot be related to his fundamental neuropsychologic defects in the same fashion. There are other differences as well; the neuropsychologic consequences of brain injury are maximal at outset and tend to either remain static or improve, and the same is true of their directly related behavioral sequelae. In the case of psychotic manifestations, however, there is often a lag after the brain damage; they may even appear, as in the case at hand, while the patient is in fact improving neuropsychologically. Furthermore, pharmacologic remediation of the psychosis occurs in the face of persistent neuropsychological impairment. Unlike the situation with prefrontal syndromes, psychiatric morbidity and neuropsychological morbidity in post-brain-injury psychosis seem clearly to follow independent courses.

**Response to Dopamine-Blocking Agents and Etiology.** Perhaps the best clue to etiology is the striking therapeutic response to dopamine-blocking agents apparent in the case presented and in similar cases. This patient responded favorably to neuroleptics from the time of his initial psychiatric symptoms and twice decompensated when the neuroleptics were withdrawn.

A response to a neurotransmitter-blocking agent can never conclusively demonstrate that derangement of that transmitter system had been fundamentally responsible for the pathologic condition. Nevertheless, there is evidence to suggest that lesions in specific brain regions, among them the inferior parietal area involved in this patient, may evoke alterations in subcortical dopaminergic regulation. The phenomenon is particularly well documented in the rat, where prefrontal lesions cause increased dopamine turnover and increased dopamine receptor number (4). There is evidence as well from human beings studied in the wake of cerebral infarction that lesions of the frontal and the parietal lobes can induce striatal hyperdopaminergia (5). Strikingly, the time course of this effect is similar to that seen in post-brain-injury psychosis, and is opposite to that encountered in postinjury neuropsychological defect; that is, it becomes more rather than less apparent as time elapses.

The brain tissue damaged in this case is classified as supramodal cortex (6) in that it receives inputs from many cortical areas for high-level sensory processing. Such areas exhibit connectivity to other regions implicated in affective and motivational behavior, including limbic and prefrontal sites. Of particular relevance to the hypothesis put forward above, the brain region damaged in the patient under discussion manifests connections to subcortical dopamine-innervated limbic sites: the nucleus accumbens exhibits connections to the inferior parietal lobule via the substantia innominata (7).

The proposed explanation of the patient's psychotic disorder, then, is that his focal brain lesion resulted in subsequent deregulation of subcortical dopaminergic activity, which in turn manifested as psychosis. Both the neuropsychologic defects and the dopaminergic derangement were induced by the lesion, but the neuropsychologic impairments are not seen as themselves responsible for inducing the psychosis.

These speculations on the relationship between brain injury and psychosis are capable of confirmation through the study of dopamine metabolism, via such techniques as positron emission tomography scanning. Cases of this sort, which so closely resemble idiopathic psychotic conditions, underscore the difficulty of relying too heavily on overt symptoms to define either a presumed etiology or a homogeneous sample of a particular psychiatric disorder.

## References

1. Davison K, Bagley CR. Schizophrenia-like psychoses associated with organic disorders of the central nervous system: a review of the literature. Br J Psychiatry 1969; (Suppl 4):113–184.
2. Achte KA, Hillbom E, Aalberg V. Psychosis following war injuries. Acta Psychiatri Scand 1969; 45:1–18.
3. Fuster JM. The prefrontal cortex; anatomy, physiology, and neuropsychology of the frontal lobes. New York: Raven Press, 1989.
4. Pycock CJ, Carter CJ, Kerwin RW. Effect of 6-hydroxydopamine lesions of the medial prefrontal cortex on neurotransmitter systems in subcortical sites in the rat. J Neurochem 1980; 34:91–99.
5. Bracha HS, Lyden PD, Khansarinia S. Delayed emergence of striatal hyperdopaminergic activity after anterolateral ischemic lesions in humans: evidence from turning behavior. Biol Psychiatry 1989; 25:265–274.
6. Mesulam MM. Principles of behavioral neurology. Philadelphia: FA Davis, 1985:25–27.
7. Mesulam MM, Geschwind N. On the possible role of neocortex and its limbic connections in the process of attention and schizophrenia: clinical cases of inattention in man and experimental anatomy in monkey. J Psychiatr Res 1978; 14:249–259.

# Short-Term and Long-Term Possibilities for Interventive Gerontology

Gregory M. Fahy, Ph.D.

---

Aging is a process that affects each of us, not only as individuals but as members of society. In the relatively near future, current trends in the demography of aging will become an increasingly urgent issue for social policy makers. In the past 85 years or so, life expectancy has risen by more than 25 years in the industrialized nations while birth rates have steadily fallen (1–3). In the United States, the number of individuals over the age of 65 first exceeded the number of teenagers in 1983 (1). Those over 85 have been the fastest-growing age group in the population for some years (1), and the unprecedented aging of populations throughout the world is creating growing economic, social, and medical pressures (1–3). It has been eloquently argued that an important part of the solution to these problems must come from direct intervention into aging itself (1, 4).

The thesis developed in this paper is that the genetic origin of most aging processes opens up real prospects for significant intervention into aging. The discussion therefore begins with a brief review of evidence for genetic control of aging and against theories of wear and tear as the primary basis of aging. This review is followed by several striking examples of the reversal of age-related changes that may have therapeutic implications. A final section considers some of the more distant possibilities for intervention, in particular the therapeutic potential of highly advanced artificial molecular constructs (5, 6).

## Aging: A Surprisingly "Simple," Malleable Process

**General Observations on Aging.** It appears likely that the complexity and lethality of aging are primarily the result of a relatively small number of "programmed" changes in genetic expression. Detailed examination of aging has revealed that age-related biochemical changes tend to be highly selective and reproducible, not global and randomized, as might be expected for wear and tear (7, 8). Furthermore, these biochemical changes, with few exceptions, characteristically occur most robustly in pivotal regulatory processes rather than in constitutively expressed genes that are under little physiologic control (7). Only a few examples in support of these general statements can be provided within the scope of this short overview. Additional and extensive supporting documentation can be found in a number of outstanding reviews (7–10).

**Evidence at Organism Level.** Genetic control of aging is implied by the wide disparity of lifespans among related animal species, including mammals (11). Also compelling is the observation that the onset of aging can be significantly (12) or even permanently (13) postponed in certain organisms by blocking development before sexual maturation.

These examples imply that senescence is dependent on the spectrum of genetic expression present in the adult organism and therefore cannot be a result just of wear-and-tear processes inherent in standard "housekeeping" metabolism or background mutations dependent only on time. Similarly, it has been found that calorie restriction extends lifespan most profoundly when it is

---

Adapted from presentations given at the First Foresight Conference on Nanotechnology, sponsored by the Stanford University Dept. of Computer Science, Palo Alto, California, Oct. 28, 1989, and the Fifth Nanotechnology Symposium of the MIT Nanotechnology Study Group, MIT, Jan. 31, 1990.

From the Transplantation Laboratory, American Red Cross Biomedical Research and Development, Rockville, Maryland. Address reprint requests to Gregory M. Fahy, Ph.D., Transplantation Laboratory, W124, American Red Cross/Jerome Holland Laboratory, 15601 Crabbs Branch Way, Rockville, MD 20855.

applied early in life, whereas temperature reduction extends life most effectively in fish when it is applied only after maturation (14); these two effects may be additive (14, 15). After puberty, lifespan may be at least doubled or trebled in many species, including one marsupial, by removing specific glands or preventing reproduction (7). Some complex organisms, most notably sea anemones (7, 16) and a variety of plants (7, 17), evidently do not age at all, the plants lasting at least many thousands of years. These organisms clearly demonstrate that aging need not inevitably result from wear and tear.

**Evidence from Studies of Gene Expression.** Such observations are supported by more specific examinations of genetic expression and gene stability during aging. Despite extensive searches, abnormal proteins resulting from somatic mutations have not been found to appear with aging (8–12). Instead, direct measurements have shown little change in the expression of most proteins and RNA transcripts with aging, but drastic changes in a handful of proteins, some of which increase and some of which decrease in concentration with age in an apparently regulated manner (for example, 18, 19, 22–25, 25a).

One major example of these general observations is provided by hormone and neurotransmitter receptors (26): genomic stability is implied by the fact that aging does not alter the binding characteristics of the great majority of receptors, whereas systematic and drastic age-related reductions in receptor populations strongly implicate physiologic control of aging.

A second striking example of genetic regulation of aging is provided by elongation factor 1 (EF1), which catalyzes the rate-limiting step of protein synthesis and whose importance is indicated by the fact that it represents 3%–11% of the total protein of most young cells (24). Although approximately 50 factors are required for synthesis of proteins from available mRNA, 30%–80% of the age-related global decrease (23, 24, 27) in protein synthetic rate appears to be due to the specific repression of EF1 (24, 25).

**Physiologic Evidence: Ready Reversibility.** Specific repression of EF1, together with several observations of the reversibility of age-related changes, including the reversibility of age-related decrements in receptor number (28), suggests the possibility of profound causal interrelationships between many aging processes, such that the correction of one defect may lead to improvements in other apparently unrelated aspects of aging.

For example, neonatal thymus transplants can reverse the age-related loss of beta receptors in the brain (29), reverse the age-related rise in the number of hepatic tetrapoid cells (30, 31), lower serum insulin levels to youthful values (32), elevate depressed plasma $T_3$ levels to values nearer those of young animals (31, 32), and restore the ability of submandibular glands to synthesize DNA in response to isoproterenol stimulation (32). Furthermore, a stable LHRH antagonist was able to induce complete regeneration of the almost totally atrophied thymus gland of old rats (33).

All these observations clearly seem to involve physiologic controls over aging and not wear-and-tear mechanisms. They also suggest a surprising underlying simplicity, since the reversal of age-related changes in response to simple physiologic stimuli demonstrates that the reversible changes are not themselves primary aging changes, but instead are secondary effects of other age-related changes.

**Hereditary Evidence.** It appears that the number of genes governing human longevity must be a tiny fraction of the total (34). This idea is consistent with the fact that relatively small changes in the genome, including the single-gene mutation of progeria (35, 36) and the simple duplication of the 21q22 region of chromosome 21 in Down's syndrome (37, 38), seem able to greatly accelerate the rate of aging in humans, while deletion of genetic information from the Y chromosome can evidently extend the mean lifespan of human males by 12–13 years (39).

**Random Damage as Effect, Not Cause.** It is true that some types of random damage accumulate with age. However, available examples of random damage seem to constitute the exceptions that prove the rule of genetic control. At least three general categories of random damage can be identified.

The first consists of slow nonenzymatic formation and accumulation of adducts on long-lived structural proteins such as lens proteins (40) and changes such as racemization of aspartic acid residues of tooth enamel protein (41). Such alterations particularly affect body components that have little capacity for self-repair, yet may in fact result from regulated decreases in protein synthesis and turnover, including reduced production of proteins needed for antioxidant defenses and other types of housekeeping (42).

Such changes are closely related to the second category of random damage: the accumulation of altered enzymes with aging (21, 43–46). These do not arise from genomic damage (21, 44–46), but instead from reduced protein turnover. Reduced turnover is in turn apparently due to compensatory decreases in protein (44, 47, 48) and mRNA catabolism that permit maintenance of overall mRNA levels (21, 23, 25) and enzyme activities (22) in the face of large global drops in both mRNA (23, 25, 49) and protein synthesis. Random damage secondary to regulated decreases in protein turnover or gene expression is not a cause, but is instead an effect of aging.

The final category of random damage is represented by the very recently described (50) age-related accumulation of DNA damage (8-hydroxy-2′-deoxyguanosine [HDG]) in rat liver, kidney, and intestine (but not in brain or testes). Given that only a 2% drop in expression of the DNA repair enzymes for HDG (an endonuclease and glycosylase) would be required to explain the observed HDG accumulation (50), and given that the endonuclease repair rate fell two- to threefold between 4 months and 12–24 months (50), an atypically large drop for proteins in general, the authors' conclusion that HDG accumulation might be "principally due to the slow loss of DNA nuclease activity" (rather than to an inherently uncontrollable random damage mechanism) would seem to be well-supported. Furthermore, the accumulation observed is not obviously sufficient to account for any age-related phenomena and could pertain primarily or exclusively to nontranscribed DNA (50).

There is no persuasive evidence that age-related changes in genetic expression are generally the result of wear-and-tear processes such as damage to DNA, and much evidence, including the evidence cited above and the examples of intervention cited below, strongly suggests the contrary.

## Molecular Approaches to Intervention: Some Existing Examples

The following selected examples of specific possibilities for intervention have been chosen to further illustrate basic aspects of aging and emphasize the range of interventive possibilities. It should, of course, be understood that properly designed clinical trials and many years of research will be required to adequately establish the efficacy and safety of any of the following possible routes to intervention.

The examples below indicate that aging can be modified significantly by the simple expedient of correcting age-related alterations in the levels of specific molecules or elements, or through pharmacologic modification of specific regulatory processes that govern important aspects of aging. For the purpose of organizing the discussion, these potential molecular intervention strategies are classified as immunologic, cellular, and empirical. Due in part to the interrelationships between aging changes, some degree of overlap between these categories is unavoidable.

### Immunologic Reconstitution

Thymic involution is the primary cause of immunologic senescence (51, 52), and immunologic senescence is probably responsible for a great deal of the present morbidity and mortality of aging, including the accelerating incidence of cancer and certain kinds of cardiovascular disease with age (51). Although many age-impaired immune functions can be reconstituted by replacement of missing thymic hormones (51–55), it is now possible to actually regenerate the thymus gland in animals, and evidence from human studies is consistent with these findings.

**Thyroid Hormone.** An extensive literature describes the effect of thyroid hormone deficiencies and excesses on thymic involution and thymic function in both animals and humans. It has long been known that patients with hyperthyroidism show little or no thymic involution (56) and little or no decline in circulating thymulin levels (57, 58) even at ages of 50–80 years, in marked contrast to controls (57, 58). Antithyroid treatment of such patients induces thymic involution (56) and reduces circulating thymic hormone levels (57), whereas the administration of L-thyroxine to hypothyroid patients (57) and to normal old mice (59) restores circulating thymulin levels to more youthful values, improves many measures of immune system function (59, 60), and, at least in animals, reverses age-related thymic involution (31). These results raise the possibility of a role for thyroid hormones in the treatment of immunologic aging.

The effects of thyroid hormone are probably due both to a direct action on the thymus (58) and to the ability of thyroid hormone to activate transcription of the growth hormone gene (61, 62).

**Growth Hormone.** The stimulatory actions of growth hormone (on muscle development; synthesis of protein, RNA, and DNA; cell division; skin wound and bone fracture healing; fat catabolism; and growth of the liver, kidneys, bones, pancreas, stomach, small intestine, spleen, and thymus gland) (62–65) are generally contrary to those of aging, and the loss of growth hormone may therefore explain many age-related changes (9, 65, 66).

Indirect support for this idea is provided by the fact that calorie restriction both extends life and elevates growth hormone levels in most species (67) (though not in the rat), and growth hormone has been called both a "hormone of fasting" and a "calorie homeostasis" hormone (63).

More direct support includes observations such as the ability of growth hormone given to 19–21-month-old rats to more than restore the protein synthetic rate of diaphragm muscle to the levels observed in 3–4-month-old rats (68). Unfortunately, attempts thus far to modify rodent lifespan by administration of growth hormone have employed doses of hormone about 100–1000 times the human dose and, perhaps for this reason, have not been successful (69, 70).

Direct support for the importance of growth hormone in human senescence has come from a number of recent clinical studies showing that growth hormone administration to elderly individuals can increase lean body mass, decrease adiposity, improve bone density, thicken skin (71, 72), elevate somatomedin C levels, lower plasma cholesterol, elevate both osteocalcin and parathyroid hormone (suggesting an increased turnover of bone), and improve nitrogen retention, renal sodium reabsorption, and renal conversion of inactive to active vitamin D (73, 74).

Although these studies have emphasized the potentially deleterious side effects of growth hormone—induction of insulin resistance, water retention, elevated triglycerides, hypertension, and cardiovascular morbidity—these effects appeared minimal in comparison to the therapeutic effects observed, particularly at lower doses, which were nevertheless fully sufficient to produce desired results (71–74). The most worrisome finding was that of Marcus et al. (73), who reported highly elevated insulin levels in response to even the lowest dose of growth hormone administered; clearly more research is necessary to evaluate the consequences and avoidability of this effect.

The possible immunotherapeutic benefits of growth hormone received impressive support from work by Kelley et al., who transplanted growth hormone-producing $GH_3$ pituitary adenoma cells into aged female Wistar-Furth rats and observed complete reconstitution of immunologic function as well as histologically complete regeneration of the thymus gland two months after transplantation into 18-month-old rats (64). Recipients 24 months old did not show complete reconstitution, but thymus weight increased and thymic histology improved dramatically; PHA-induced T-cell proliferation increased 6.3-fold; and interleukin 2 production increased 17-fold (75). Longer treatment times were not investigated. Follow-up studies showed that direct administration of exogenous growth hormone improves spleen cell proliferative response to Con A and PHA, elevates natural killer cell activity in 26-month-old Fisher 344 rats (76), and elevates plasma thymulin levels in aged dogs (77). In women aged 56.3 ± 8.7 years who showed impaired endogenous growth hormone secretion, exogenous growth hormone nearly doubled the "specific lytic activity" of natural killer (NK) cells, and somatomedin C levels correlated with NK activity (72).

The ultimate reason for the decline in growth hormone secretion with aging may be depletion of dopamine and norepinephrine from the hypothalamus rather than any serious failure of capacity for growth hormone synthesis or release (7, 9). In humans aged 63–78, growth hormone release in response to GHRH is sluggish but can be restored to about 80% of the release response of 18–31-year-olds by priming the elderly individuals with GHRH every other day for 12 days (78).

**Arginine.** Arginine is known to be an effective stimulator of growth hormone release from the pituitary (for review, see ref. 79). Because of this and more direct actions, arginine has many useful and relevant biologic effects. For example, old rats given arginine-supplemented diets had accelerated rates of skin wound healing (80). More pertinent to immune system reconstitution, children (81), adults (82), and elderly (63–88-year-old) individuals (83) have all shown significantly enhanced immune system function in response to arginine administration.

These observations are consistent with the finding in animals that arginine administration results in enhanced thymus weight, enhanced mitogen response, reduced thymic involution response to a variety of stresses, enhanced skin allograft rejection, enhanced immune response to chemically induced tumors, and enhanced recognition and rejection of transplanted tumors (79).

Arginine is evidently nontoxic to humans in high doses. However, arginine is a secretagogue for prolactin, insulin, glucagon, somatostatin, pancreatic polypeptide, and adrenal catecholamines, and use in patients with renal and hepatic insufficiency may lead to dangerous hyperkalemia (79). Any therapeutic use of arginine for the treatment of aging must take these additional effects into account.

Ornithine is a more potent and longer-lasting inducer of pituitary growth hormone release than arginine and does not promote the release of insulin (79). It shares arginine's growth-hormone-releasing, wound-healing, and thymotropic properties.

Zinc. Zinc is required for the activity of over 200 cellular enzymes, including RNA polymerases, DNA polymerases, thymidine and adenylate kinase, and superoxide dismutase, and it is also required for nerve growth factor activity (84). Zinc is necessary for synthesis or degradation of DNA, RNA, protein, lipid, and carbohydrates (84). That simple dietary supplementation with zinc could be a surprisingly powerful, inexpensive, and safe means of intervening in immunologic and other aspects of aging is indicated by a number of observations.

For example, aging leads, in both humans (85) and animals (86), to a quantitatively small, but functionally very important, deficiency of zinc. Like aging, zinc deficiencies are known to reduce circulating levels of thyroid hormones ($T_3$ and $T_4$), growth hormone, and thymic hormones and produce many signs of immunologic deficiency (84). "Facteur thymique serique" can compete with thymulin for binding to thymulin receptors but is inactive without zinc; zinc administration converts this factor into active thymulin and reverses the problem (85). Zinc deficiencies also elevate the "geromimetic" hormones cortisol and corticosterone (84).

Zinc administration to elderly humans has been shown to increase skin reactions to local antigen injections (87, 88), a test that correlates with human life expectancy (88, 89); to elevate antibody responses to tetanus toxoid (87); to elevate levels of circulating T cells (87); and to completely reverse the age-related decline in circulating active and total thymulin levels (85, 90). Zinc administration to rodents prevents age-related changes in thymic hormone production (91), T-cell activities (91), natural killer cell cytotoxicity (91), and autoimmunity (92) and reverses age-related depression of $T_3$ levels (93) as well as reversing histologic evidence of thymic involution (84).

### Cellular Strategies

"Cellular" interventions as discussed here are strategies for maintaining the fundamental "housekeeping" processes of cells.

**Centrophenoxine.** As noted above, a serious problem of aging is the genetically "programmed" drop in transcription of EF-1 in both mice (94, 95) and *Drosophila* (96). The value of preventing this age-related change is suggested by recent experiments (97) with transgenic *Drosophila* prepared with an extra F1 ("housekeeping") copy of the EF1-alpha gene under the control of the promoter and termination sequences of the heat shock protein hsp70; unlike the control regions for EF1 itself, these sequences are not known to vary in expression with age. Moreover, the expression of the extra EF1 inserted in this fashion is temperature-dependent, being greater at 29.5°C than at 25°C. It was found that *Drosophila* possessing augmented quantities of EF1 have mean lifespans that are 18% longer ($p < 0.001$) at 25°C and 41% longer ($p \ll 0.001$) at 29.5°C. In fact, at the higher temperature, the mean lifespan of the EF1-supplemented flies was virtually equivalent to the maximum lifespan of control *Drosophila* (97). Thus, maintaining protein synthesis appears important for longevity.

The possibility of normalizing the turnover of both RNA and protein in the intact mammal through pharmacologic means has been clearly demonstrated. Centrophenoxine stimulates mRNA and total RNA synthesis in the aged mammalian brain and liver in vivo (49), stimulates RNA and protein synthesis in human glial cells in stationary culture in vitro (98), and reduces the rate of accumulation of lipofuscin (an intracellular fluorescent material postulated by some to form as a result of a reduced rate of protein catabolism [99]) in nerve cells (100, 101).

The ability of centrophenoxine to reverse the global age-related decline in mRNA synthesis in vivo has been confirmed using an in vitro chromatin model system (25), and preliminary unpublished results indicate that centrophenoxine also restores the transcription of EF1 genes in the same model system (Webster, personal communication). It is not known whether these effects could be due to an ability of centrophenoxine to masquerade as or stimulate the production of somatomedin C.

In keeping with these effects on nervous tissue, centrophenoxine has been shown to reverse the age-related drop in the surface density of cerebellar synapses (102), restore cerebral beta adrenoceptors (103), and improve learning in mice (104). Centrophenoxine added to the drinking water of old male mice significantly increased both mean and maximum life expectancy (105), and centrophenoxine was also found to extend the lifespan of human lung cells maintained in culture (106).

Centrophenoxine is widely used in Europe as a treatment for geriatric cognitive impairment. However, a major drawback of centrophenoxine is its rapid hydrolysis in the body. Webster found that both the restoration of global mRNA synthesis (25) and the restoration of EF1 mRNA synthesis (personal communication) by centrophenoxine depend on the intact centrophenoxine molecule, its hydrolysis products being ineffective. A stable version of centrophenoxine that lacks the labile ester linkage could have significantly enhanced therapeutic effects.

**DNA Polymerase Controlling.** DNA polymerase alpha (DPA) is necessary both for the DNA synthesis required for cell division and for excision repair of DNA (107, 108). Intriguingly, fibroblast DPA exists as two isozymes, a "youthful" isozyme (or A2 form) and a "senescent" isozyme (or A1 form) (108, 109). The latter binds DNA much more poorly, adds bases much more slowly, and makes many more errors than the "youthful" or A2 form of DPA (108–110). With age, the ratio of A2 to A1 DPA declines drastically (108, 109). However, the properties of senescent DPA can be restored fully to those of the youthful form by addition of inositol-1,4-bisphosphate (108, 109), a product of the phosphatidylinositol second-messenger cascade.

The latter finding suggests that part of the problem with cell division during aging is loss of the necessary hormonal signals required for maintaining adequate rates of cell division and DNA repair (65, 66, 109). This idea is consistent with the restoration of submandibular gland DNA synthesis by thymus transplants, mentioned above (32), and with the fivefold reduction in mitotic index in liver, spleen, stomach, small intestine, and pancreas in response to hypophysectomy and its reversal by growth hormone (65).

**Restoring Cellular Proliferation.** Studies of the control of cellular proliferation are proceeding rapidly, both within (111) and outside (112) the context of aging. Cells growing in vitro gradually lose the capacity to proliferate and may eventually die (so-called "in vitro aging" or "replicative senescence"). The possible relevance of replicative senescence in vitro to aging in vivo has been discussed by Gelfant and Smith (113) and is supported indirectly by a number of biochemical parallels between altered gene expression with age in vivo and in vitro (West, personal communication) and by slowed cell division in vivo (65).

Several recent studies have been suggestive of a coming ability to prevent or reverse the senescent decline in cellular proliferative competence. A few examples: the ability of cells containing high levels of active DPA, but not cells containing lower DPA levels, to restore DNA synthesis in "senescent" cell nuclei (114) after fusion with replicatively senescent cells; the detection of high-abundance antiproliferative mRNA produced only in senescent cells that may code for a membrane-associated protein that inhibits DNA synthesis (115); the assignment of proliferative senescence genes to chromosome 1 (116); the observation of an escape from "in vitro aging" by altered culture conditions (117); and most recently, the use of antisense oligonucleotides to specifically block expression of interleukin 1-alpha and thereby more than double the number of population doublings human endothelial cells are capable of in vitro (118). The recent introduction of retin A is a notable practical example of an improved ability to control, maintain, and restore proliferative capacity, achieved in an extremely simple manner.

West has recently discovered that replicative senescence of a variety of cells (fibroblasts, endothelial cells, astrocytes, and chondrocytes) is under the control of a "senescence response element" or sen locus, and that the senescent phenotype of these cells in culture, and of skin cells in vivo, can be induced and reversed at will, making it possible to immortalize these cells without transformation (West, personal communication; 118a).

## Empirical Strategies

Empirical strategies are interventions that appear effective, but whose fundamental basis is still speculative or unknown.

**Coenzyme $Q_{10}$.** Coenzyme $Q_{10}$ (ubiquinone) is essential for normal mitochondrial ATP production (119), and its concentrations are highest in cardiac muscle (120). It is therefore of considerable interest that cardiac levels of coenzyme $Q_{10}$ drop by over 60% in rats just before the age of rapidly accelerating mortality (120), and that serious losses of myocardial coenzyme $Q_{10}$ are also associated with age-related heart disease in humans (121). Mice given coenzyme $Q_{10}$ at a dose roughly equivalent to one-third of the standard human pharmaceutical dose of 30 mg/day had mean lifespans 12 ± 0.5% longer than controls and life expectancies 56% longer than controls; maximum lifespan was extended by 42 ± 2%, maximum life expectancy by more than 100% (122). These results were obtained without producing weight loss (that is, they were not due to self-imposed calorie restriction). A lifespan experiment currently in progress at UCLA involving coenzyme $Q_{10}$ (given in the diet at a level of 0.1% by weight) is thus far confirming both the life-extending effect of coenzyme $Q_{10}$ and its lack of appetite suppression (Harris, personal communication).

Coenzyme $Q_{10}$ has been in use as a pharmaceutical in Japan since 1965 (123) and is currently consumed by several million Japanese annually. It is used in Japan for the treatment of many of the diseases of aging: cardiac arrhythmias, hypertension (both systolic and diastolic), congestive heart failure, diabetes, and periodontal disease (123). At least seven books are now available describing the biomedical and clinical applications of coenzyme $Q_{10}$ (119, 124–129). In 15 studies carried out from 1967 through 1975, some of them double blind, an average of 77% of the 285 patients treated for cardiovascular diseases and diabetes showed at least some benefit (123). Documentation of the protective effects of coenzyme $Q_{10}$ against transient ischemia is overwhelming (e.g., 119, 123–130). Coenzyme $Q_{10}$ has been reported to be immunostimulatory in both humans (131) and animals (122), a single injection partially reversing age-related immunologic decline in the latter.

Although the mechanism(s) of action of coenzyme $Q_{10}$ remain to be established, its role as an antioxidant may be dominant (130). Orally administered, it is reduced to ubiquinol, a compound recently identified as a major natural antioxidant in humans (130). Coenzyme $Q_{10}$ has been shown to reduce peroxidized mitochondrial lipid levels in the normal rat liver and heart in vivo by one-third (132), to block free radical damage secondary to carbon tetrachloride administration (132), and to block the destructive action of phospholipase $A_2$ and phospholipase C on membranes (132). Coenzyme $Q_{10}$ may thus help to counteract loss of enzymatic antioxidant defenses with aging.

Coenzyme $Q_{10}$ has no reported significant side effects, toxicity, teratogenicity, or contraindications. A proposed scenario for coenzyme $Q_{10}$–induced free radical pathology after ischemia proposed by Demopoulous and Flamm (133) is contradicted by all the studies on ischemia carried out to date.

**Deprenyl.** Deprenyl was introduced in the United States in 1989 for slowing the progression of Parkinson's disease and prolonging the effectiveness of levodopa therapy of parkinsonism. Its properties have been reviewed in detail elsewhere (134), including a recent review in this journal (135). Several investigators now view deprenyl as an intriguing substance for modifying the aging process.

Parkinson's disease is caused by the selective loss of 70% (136) to 80% (137) of the dopamine from the human striatum due to an acceleration of the normal, cardinal age-related process of selective striatal dopamine loss (138). This loss amounts to a 13% reduction per decade after the age of 45 in normal individuals (135) and may reflect a loss of dopaminergic neurons from the substantia nigra (137). Recent prospective double-blind clinical trials on parkinsonian patients have shown that deprenyl delays the need for levodopa therapy and slows the progression of the disease (139, 140), confirming and extending previous retrospective studies (141, 142). These findings suggest that deprenyl might also slow the normal aging of the human striatum which, even if not leading to frank Parkinson's disease, might well lead to less robust motor control.

Many observations suggest a general relevance of deprenyl to interventive gerontology. Deprenyl is known to greatly elevate both brain dopamine levels and brain dopamine turnover (135). This might be expected to simulate the effects of exogenous L-DOPA administration. It has been reported that L-DOPA administration releases growth hormone in both animals (143) and humans (144, 145), even resulting in increased muscle protein synthesis secondary to this release in animals (143). Dietary L-DOPA has also resulted in sizable and dose-dependent extensions of lifespan in Swiss Albino mice (146, 147), a finding I have confirmed using C57B1/6J mice (Fahy, unpublished results). Deprenyl also appears to have some effectiveness in the treatment of Alzheimer's disease (148). Finally, direct evidence for a pivotal role in aging comes from reports by Knoll of dramatic life-extending and performance-enhancing effects of deprenyl itself when given to elderly CFY or Wistar/Logan rats (135, 136, 149). For example, deprenyl elevated mean lifespan by 31%–41%, maximum lifespan by 38%–48%, mean life expectancy at the beginning of treatment by 105%–129%, and maximum life expectancy by 103%–142% (136). The minimum life expectancies of deprenyl-treated rats exceeded the maximum life expectancies of matched control rats by 23%–57%. Knoll's results appear to be at least partly confirmed by studies from at least two independent laboratories (150; Ingram, personal communication).

The mode of action of deprenyl in extending lifespan remains speculative. Deprenyl's ability to act as an effective and safe monoamine oxidase B (MAOB) inhibitor and to block neuronal uptake of certain toxins is thought to reduce free radical damage to dopaminergic cells (134, 135, 137, 151). If this is correct, even this oxidative mode of death could be physiologically regulated in that brain MAOB activity consistently rises with aging in both animals and humans, increasing by about 76% in human hindbrain between the ages of 35 and 70+ (152). Given that MAOB is made by glial cells (135) and that gliosis of the aging brain is believed to arise from the impact of peripheral hormones (7, 9, 153), an extrinsic origin of this change is quite possible.

Whatever its mode of action, deprenyl clearly represents a powerful and exciting tool not only for modifying aging, but also for probing its mechanisms. As such, it provides a good final illustration of the potential of presently available opportunities for interventive gerontology.

### Molecular Approaches to Intervention: Future Possibilities

The total budget of the National Institute on Aging for 1991 is expected to be about $325 million (154). The private sector, however, now spends about $3.6 billion annually to develop pharmaceuticals that can be used for the treatment of aging (154). This vast effort to synthesize novel therapeutic molecules can be expected to lead to increasingly sophisticated and complex molecular constructs. Moreover, the drive to develop ever more sophisticated molecular technologies is fueled also by pressures in many other industries, including the chemical, computer, and environmental industries (6).

**Molecular Devices.** The logical result of these pressures is an eventual ability to manufacture not just single-function pharmaceuticals, but actual molecular devices (5, 6). Complex, versatile molecular devices, in turn, might have interesting applications for interventive gerontology (6). For example, sufficiently sophisticated molecular devices might be able to gain access to and repair age-related changes in structures such as the lens or myelin that simple pharmaceuticals could not address. In the longer run, they could also be instrumental in imposing permanent patterns of genetic expression that block aging processes.

**Construction Pathways.** Simple molecular "devices" have been synthesized using a variety of creative approaches. These include the reagents of organic chemistry, "abzymes" (specific catalytically active antibodies) (155), "cavitands" or "chemzymes" (artificial and compact yet functionally enzymelike molecules) (156), and the innumerable constructs of modern molecular biology. Simply combining such constructs could result in more versatile and multifunctional structures. Employing emerging new tools for direct molecule-by-molecule manipulation based on the scanning tunneling microscope could further widen the range of opportunities (157–159).

The design of truly powerful molecular devices might best proceed from the present ability to synthesize any desired protein using the methods of today's immense and growing genetic engineering industry. Living cells already synthesize complex molecular machines such as enzymes, ribosomes, and nuclear pores. The main problem for molecular engineers is not an inability to construct novel proteins, but rather the inability to know which novel proteins to construct. However, understanding how to select amino acid sequences that will give rise to proteins that carry out desired specific novel functions is of intense current research interest (160–162) and major economic import. Several books are available on the emerging discipline of protein engineering (160, 161), and at least one catalytically active novel protein has in fact been designed from first principles, constructed, and shown to function as designed (162).

But protein engineering does not require design from first principles. Reasonable first approximations of the "correct" amino acid sequence for a given task can be obtained from the wealth of present knowledge in protein folding, biochemistry, and organic chemistry (5, 6, 160–162). The final sequence can then be found by an iterative process in which proteins with several likely sequences are identified, prepared, examined, and the results used to modify the sequences selected for the next iteration (5, 6).

Carried to its logical conclusion, this process could permit synthesis of just about any desired molecular assemblage. For example, an early step could be the synthesis of an enzyme that catalyzes the linkage of completely nonbiologic functional groups onto other enzymes or structural proteins. These in turn could be designed to self-assemble onto molecular scaffolds containing binding sites for other catalytically active molecules, and so on.

**Possible Results.** Basic designs exist for molecular-scale mechanical computers that could have memory capacities of $10^9$ bytes per cubic micron, gigahertz clock speeds, error rates of less than one per $10^{12}$ operations, and heat dissipations of about $4 \times 10^{-22}$ cal per logic operation (6, 163, 164). This degree of raw power is surely far more than would be required for repairing accumulated damage to nonrepaired structures during aging; for most such tasks, relatively "dumb" molecular repair devices not requiring molecular computers ought to be sufficient. But such potential power does lend support to the idea that relatively complex, permanent modifications of genetic expression might be attainable that could block otherwise-programmed aging processes.

The principles are sufficiently clear to warrant confidence that complex therapeutic molecular devices can be constructed, and the examples available from biology demonstrate the spectacular range of possibilities already exploited by living systems. However, it is obvious that the full flowering of molecular engineering technology will take a long time. In the meantime, increasing success in controlling regulated features of aging by more conventional means should permit many important goals to be attained.

## Summary

Worldwide demographic trends, including the aging of the human population and the steadily declining fertility rate in developed nations, are creating enormous economic pressures due to the ever-increasing demand for health care services for the elderly and the ever-decreasing ability of the young to pay for these services. An attractive cost-containment strategy is accelerated research on basic molecular mechanisms of aging and rapid clinical application of the results. This approach should result in maintenance of health and productivity over a longer fraction of the lifespan and thereby to a reduction in the ratio of health care expenditures to lifelong earnings. This strategy should lead to many important improvements in the human condition. Sooner or later, our armamentarium will be supplemented by powerful and useful tools from the field of molecular engineering. With proper care, both these developments may help us to live longer, healthier, happier lives.

### References

1. Butler RN. A changing demography. In: Butler RN, Bearn AG, eds. The aging process: therapeutic implications. New York: Raven Press, 1985:5–19.
2. Longman P. Born to pay: the new politics of aging in America. Boston: Houghton Mifflin, 1987.
3. Wattenburg BJ. The birth dearth. New York: Pharos Books, 1987.
4. Anonymous. The research gap: the need for boosting research to achieve independence for older Americans. Washington, DC: Alliance for Aging Research (2021 K St., NW, Suite 305, Washington, DC 20036), 1990.
5. Drexler KE. Molecular engineering: an approach to the development of general capabilities for molecular manipulation. Proc Natl Acad Sci USA 1981; 78:5275–5278.
6. Drexler KE. Engines of creation. Garden City, NY: Anchor Press/Doubleday, 1986.
7. Finch CE. Neural and endocrine determinants of senescence: investigation of causality and reversibility by laboratory and clinical interventions. In: Warner HR et al., eds. Modern biological theories of aging. New York: Raven Press, 1987:261–308.
8. Finch CE. The regulation of physiological changes during mammalian aging. Q Rev Biol 1976;51:49–83.
9. Meites J. Effects of aging on the hypothalamic-pituitary axis. Rev Biol Res Aging 1990; 4:253–261.
10. Everitt AV, Burgess JA. Hypothalamus pituitary and aging. Springfield, IL: Charles C Thomas, 1976.
11. Cutler RG. Evolutionary biology of aging and longevity in mammalian species. In: Johnson JE Jr, ed. Aging and cell function. New York: Plenum Press, 1984:1–145.
12. Miller SE, Hadfield MG. Developmental arrest during larval life and lifespan extension in a marine mollusc. Science 1990; 248:356–358.
13. Beck SD, Bharadwaj RK. Reversed development and cellular aging in an insect. Science 1972; 178:1210–1211.
14. Liu RK, Walford RL. Mid-life temperature-transfer effects on life-span of annual fish. J Gerontol 1975; 30:129–131.
15. Walford R. Maximum lifespan. New York: WW Norton, 1983.
16. Comfort A. The process of aging. New York: New American Library of World Literature, 1964.
17. Touzet G. Perennial plants. Interdiscipl Topics Gerontol 1985; 21:263–283.
18. Fleming JE, Quattrocki E, Latter G, et al. Age-dependent changes in proteins of *Drosophila melanogaster*. Science 1986; 231:1157–1159.
19. Fleming JE, Melnikoff PS, Latter GJ, Chandra D, Bensch KG. Age dependent changes in the expression of *Drosophila* mitochondrial proteins. Mech Ageing Dev 1986; 34:63–72.
20. Sohal RS, Birnbaum LS, Cutler RG, eds. Molecular biology of aging, gene stability and gene expression. New York: Raven Press, 1985.
21. Gershon D. Current status of age altered enzymes: alternative mechanisms. Mech Ageing Dev 1979; 9:189–196.
22. Finch CE. Enzyme activities, gene function and ageing in mammals (review). Exp Gerontol 1972; 7:53–67.
23. Richardson A, Rutherford MS, Birchenall-Sparks MC, Roberts MS, Wu WT, Cheung HT. Levels of specific messenger RNA species as a function of age. In: Sohal RS, Birnbaum LS, Cutler RG, eds. Molecular biology of aging, gene stability and gene expression. New York: Raven Press, 1985:229–241.
24. Webster GC. Protein synthesis in aging organisms. In: Sohal RS, Birnbaum LS, Cutler RG, eds. Molecular biology of aging, gene stability and gene expression. New York: Raven Press, 1985:263–289.
25. Webster GC. Protein synthesis. In: Lints FA, Soliman MH, eds. *Drosophila* as a model organism for ageing studies. London: Blackie, 1988:119–128.

25a. Chatterjee B, Fernandes G, Yu BP, et al. Calorie restriction delays age-dependent loss in androgen responsiveness of the rat liver. FASEB J 1989; 3:169–173.

26. Roth GS. Altered hormone/neurotransmitter action during aging: receptor and post-receptor changes. In: Roy AK, Chatterjee B. Molecular basis of aging. New York: Academic Press, 1984:167–185.
27. Makrides SC. Protein synthesis and degradation during aging and senescence. Biol Rev Cambr Philos Soc 1983; 58:343–422.
28. Levin P, Haji M, Joseph JA, Roth GS. Effect of aging on prolactin regulation of rat striatal dopamine receptor concentrations. Life Sci 1983; 32:1743–1749.
29. Piantanelli L, Gentile S, Fattoretti P, Viticchi C. Thymic regulation of brain cortex beta-receptors during devel-

opment and aging. Arch Gerontol Geriatr 1985; 4:179–185.

30. Pieri G, Giuli C, Del Moro M, Piantanelli L. Electron-microscopic morphometric analysis of mouse liver. II. Effect of ageing and thymus transplantation in old animals. Mech Ageing Dev 1980; 13:275–283.
31. Fabris N, Mocchegiani E, Muzzioli M, Provinciali M. Neuroendocrine-thymus interactions: perspectives for intervention in aging. Ann New York Acad Sci 1988; 521:72–87.
32. Piantanelli L, Basso A, Muzzioli M, Fabris N. Thymus-dependent reversibility of physiological and isoproterenol evoked age-related parameters in athymic (nude) and old normal mice. Mech Ageing Dev 1978; 7:171–182.
33. Greenstein BD, Fitzpatrick FTA, Kendall MD, Wheeler MJ. Regeneration of the thymus in old male rats treated with a stable analogue of LHRH. J Endocrinol 1987; 112:345–350.
34. Cutler RG. Evolution of human longevity and the genetic complexity governing aging rate. Proc Natl Acad Sci USA 1975; 72:4664–4668.
35. Beauregard S, Gilchrest BA. Syndromes of premature aging. Dermatol Clin 1987; 5:109–121.
36. Badame AJ. Progeria. Arch Dermatol 1989; 125:540–544.
37. Sinex FM, Merril CR, eds. Alzheimer's disease, Down's syndrome, and aging. Ann NY Acad Sci 1982; 396:1–192.
38. Smith GF, Warren ST. The biology of Down syndrome. In: Smith GF, ed. Molecular structure of the number 21 chromosome and Down syndrome. Ann NY Acad Sci 1985; 450:1–9.
39. Fahy GM. Gender and longevity. Life Extension Rep 1990; 10:17–19.
40. Manabe S, Bucala R, Cerami A. Nonenzymatic addition of glucocorticoids to lens proteins in steroid-induced cataracts. J Clin Invest 1984; 74:1803–1810.
41. Masters-Helfman P, Bada JL. Aspartic acid racemization in tooth enamel from living humans. Proc Natl Acad Sci USA 1975; 72:2891–2894.
42. Taylor A, Davies KJ. Protein oxidation and loss of protease activity may lead to cataract formation in the aged lens. Free Radic Biol Med 1987; 3:371–377.
43. Reiss U, Gershon D. Comparison of cytoplasmic superoxide dismutase in liver, heart, and brain of aging rats and mice. Biochem Biophys Res Commun 1976; 73:255–261.
44. Rothstein M. The formation of altered enzymes in aging animals. Mech Ageing Dev 1979; 9:197–202.
45. Gershon D, Reznick A, Reiss U. Characterization and possible effects of age-associated alterations in enzymes and proteins. Aging 1979; 8:21–26.
46. Jacobus S, Gershon D. Age-related changes in inducible mouse liver enzymes: ornithine decarboxylase and tyrosine aminotransferase. Mech Ageing Dev 1980;12:311–322.
47. Reznick AZ, Gershon D. The effect of age on the protein degradation system in the nematode *Turbatrix aceti.* Mech Ageing Dev 1979; 11:403–415.
48. Lavie L, Reznick AZ, Gershon D. Decreased protein and puromycinyl-peptide degradation in livers of senescent mice. Biochem J 1982; 202:47–51.
49. Zs-Nagy I, Semsei I. Centrophenoxine increases the rates of total and mRNA synthesis in the brain cortex of old rats: an explanation of its action in terms of the membrane hypothesis of aging. Exp Gerontol 1984; 19:171–178.
50. Fraga CG, Shigenaga MK, Park J-W, Degan P, Ames BN. Oxidative damage to DNA during aging: 8-hydroxy-2′-deoxyguanosine in rat organ DNA and urine. Proc Natl Acad Sci USA 1990; 87:4533–4537.
51. Goldstein AL, Low TLK, Hall N, Naylor PH, Zatz MM. Thymosin: can it retard aging by boosting immune capacity? In: Regelson W, Sinex FM, eds. Intervention in the aging process. Part A: Quantitation, epidemiology, and clinical research. New York: Alan R Liss, 1983:169–197.
52. Weksler ME, Innes JB, Goldstein G. Immunological studies of aging. 4: The contribution of thymic involution to the immune deficiencies of aging mice and reversal with thymopoietin. J Exp Med 1978; 148:996–1006.
53. Cowan MJ, Fujiwara P, Wara DW, Amman AJ. Effect of thymosin on cellular immunity in old age. Mech Ageing Dev. 1981; 15:29–40.
54. Goldstein AL, Low TL, Thurman GB, et al. Current status of thymosin and other hormones of the thymus gland. Recent Prog Hormone Res 1981; 37:369–415.
55. Meroni PL, Barcellini W, Borghi MO, et al. Immunopotentiating activity of thymopentin treatment in elderly subjects. In: Goldstein AL, ed. Biomedical advances in aging. New York: Plenum Press, 1990:537–550.
56. Simpson JG, Gray ES, Michie W, Beck S. The influence of preoperative drug treatment on the extent of hyperplasia of the thymus in primary thyrotoxicosis. Clin Exp Immunol 1975; 22:249–255.
57. Fabris N, Mocchegiani E, Mariotti S, Pacini F, Pinchera A. Thyroid function modulates thymic endocrine activity. J Clin Endocrinol Metab 1986; 62:474–478.
58. Fabris N, Mocchegiani E, Mariotti S, Pacini F, Pinchera A. Thyroid-thymus interactions during development and aging. Hormone Res 1989; 31:85–89.
59. Fabris N, Mocchegiani E. Endocrine control of thymic serum factor production in young-adult and old mice. Cell Immunol 1985; 91:325–335.
60. Fabris N, Muzzioli M, Mocchegiani E. Recovery of age-dependent immunological deterioration in Balb/C mice by short-term treatment with L-thyroxine. Mech Ageing Dev 1982; 18:237–338.
61. Jones PM, Burrin JM, Ghatei MA, O'Halloran DJ, Legon S, Bloom SR. The influence of thyroid hormone status on the hypothalamo-hypophyseal growth hormone axis. Endocrinology 1990; 126:1374–1379.
62. Felig P, Baxter JD, Broadus AE, Frohman LA. Endocrinology and metabolism, 2nd ed. NY: McGraw Hill, 1987.
63. Brown GM, Reichlin S. Psychologic and neural regulation of growth hormone secretion. Psychosom Med 1972; 34:45–61.
64. Kelley KW, Brief S, Westly HJ, et al. $GH_3$ pituitary adenoma cells can reverse thymic aging in rats. Proc Natl Acad Sci USA 1986; 83:5663–5667.
65. Rudman D. Growth hormone, body composition, and aging. J Am Geriatr Soc 1985; 33:800–807.
66. Sonntag WE, Meites J. Decline in growth hormone secretion in aging animals and man. Interdiscipl Topics Gerontol 1988; 24:111–124.
67. Thomas GB, Mercer JE, Karalis T, Rao A, Cummins JT, Clarke IJ. Effect of restricted feeding on the concentrations of growth hormone (GH), gonadotropins, and prolactin (PRL) in plasma, and on the amounts of messenger ribonucleic acid for GH, gonadotropin subunits,

and PRL in the pituitary glands of adult ovariectomized ewes. Endocrinology 1990; 126:1361–1367.

68. Sonntag WE, Hylka VW, Meites J. Growth hormone restores protein synthesis in skeletal muscle of old male rats. J Gerontol 1985; 40:689–694.
69. Emerson JD, Emerson GM. Effect of chronic growth hormone plus prednisolone on life span of the female rat. Fed Proc 1962; 21:198.
70. Everitt AV. The effect of pituitary growth hormone on the aging male rat. J Gerontol 1959; 14:415–424.
71. Rudman D, Feller AG, Nagraj HS, et al. Effects of human growth hormone in men over 60 years old. N Engl J Med 1990; 323:1–6.
72. Crist DM, Peake GT, Mackinnon LT, Sibbitt WL Jr, Kraner JC. Exogenous growth hormone treatment alters body composition and increases natural killer cell activity in women with impaired endogenous growth hormone secretion. Metabolism 1987; 36:1115–1117.
73. Marcus R, Butterfield G, Holloway L, et al. Effects of short term administration of recombinant human growth hormone to elderly people. J Clin Endocrinol Metab 1990; 70:519–527.
74. Binnerts A, Wilson JHP, Lamberts WJ. The effects of human growth hormone administration in elderly adults with recent weight loss. J Clin Endocrinol Metab 1988; 67:1312–1316.
75. Kelley KW, Davila DR, Brief S, Simon J, Arkins S. A pituitary-thymus connection during aging. Ann NY Acad Sci 1988; 521:88–98.
76. Davila DR, Brief S, Simon J, Hammer RE, Brinster RL, Kelley KW. Role of growth hormone in regulating T-dependent immune events in aged, nude, and transgenic rodents. J Neurosci Res 1987; 18:108–116.
77. Goff BL, Roth JA, Arp LH, Incefy GS. Growth hormone treatment stimulates thymulin production in aged dogs. Clin Exp Immunol 1987; 68:580–587.
78. Iovino M, Monteleone P, Steardo L. Repetitive growth hormone-releasing hormone administration restores the attenuated growth hormone (GH) response to GH-releasing hormone testing in normal aging. J Clin Endocrinol Metab 1989; 69:910–913.
79. Barbul A. Arginine: biochemistry, physiology, and therapeutic implications. J Parenter Enter Nutr 1986; 10:227–234.
80. Chyun JH, Griminger P. Improvement of nitrogen retention by arginine and glycine supplementation and its relation to collagen synthesis in traumatized mature and aged rats. J Nutr 1984; 114:1697–1704.
81. Alexander JW, MacMillan BG, Stinnett JD, et al. Beneficial effects of aggressive protein feeding in severely burned children. Ann Surg 1980; 192:505–517.
82. Barbul A, Sisto DA, Wasserkrug HL, et al. Arginine stimulates lymphocyte immune responses in healthy humans. Surgery 1981; 90:244–251.
83. Fabris N, Mocchegiani E, Muzzioli M. Recovery of age-related decline of thymic endocrine activity and PHA response by lysin-arginine combination. Int J Immunopharmacol 1986; 8:677–685.
84. Fabris N, Mocchegiani E, Muzzioli M, Provinciali M. Zinc, immunity, and aging. In: Goldstein A, ed. Biomedical aging research. New York: Plenum Press, 1990:271–281.
85. Fabris N, Mocchegiani E, Amadio L, Zannotti M, Licastro F, Franceschi C. Thymic hormone deficiency in normal ageing and Down's syndrome: is there a primary failure of the thymus? Lancet 1984; 1:983–986.
86. Muzzioli M, Mocchegiani E, Provinciali M, Zaia AM, Palloni R, Fabris N. Recovery of age-related immune deterioration by zinc supplementation in mice. Trends Biomed Gerontol 1988; 1:173–174.
87. Ducheateu J, Pleau JM, Vrijen R, Coolet H. Beneficial effects of oral zinc supplementation on the immune response of old people. Am J Med 1981; 70:1001–1004.
88. Chandra RK. Nutritional regulation of immune function at the extremes of life: in infants and in the elderly. In: White P, Selvey N, eds. Malnutrition: determinants and consequences. New York: Alan R Liss, 1984: 245-X.
89. Wayne SJ, Rhyne RL, Garry PJ, Goodwin JS. Cell-mediated immunity as a predictor of morbidity and mortality in subjects over 60. J Gerontol 1990; 45:M45–M48.
90. Travaglini P, Moriondo P, Togni E, et al. Effect of oral zinc administration on prolactin and thymulin circulating levels in patients with chronic renal failure. J Clin Endocrinol Metab 1989; 68:186–190.
91. Iwata T, Incefy GS, Tanaka T, et al. Circulating thymic hormone levels in zinc deficiency. Cell Immunol 1979; 47:100–105.
92. Beach RS, Gershwin ME, Hurley LS. Nutritional factors and autoimmunity. I. Immunopathology of zinc deprivation in New England mice. J Immunol 1981; 126:1999–2006.
93. Mocchegiani E, Fabris N. Neuroendocrine-thymus interactions. I. "In vitro" modulation of thymic factor secretion by thyroid hormones. J Endocrinol Invest 1990; 13:139–147.
94. Blazejowski CA, Webster GC. Decreased rates of protein synthesis by cell-free preparations from different organs of aging mice. Mech Ageing Dev 1983; 21:345–356.
95. Blazejowski CA, Webster GC. Effect of age on peptide chain initiation and elongation in preparations from brain, liver, kidney and skeletal muscle of the C57Bl/6J mouse. Mech Ageing Dev 1984; 25:323–333.
96. Webster GC, Webster SL. Specific disappearance of translatable messenger RNA for elongation factor one in aging *Drosophila melanogaster*. Mech Ageing Dev 1984; 24:335–342.
97. Shepard JC, Walldorf U, Hug P, Gehring WJ. Fruit flies with additional expression of the elongation factor EF-1 alpha live longer. Proc Natl Acad Sci USA 1989; 86:7520–7521.
98. Ludwig-Festl VM, Grater B, Bayreuther K. Erhohung von Zellstoffwechselleistungen in normalen, diploiden, menschlichen Glia-Zellen in stationaren Zellkulturen induziert durch Meclofenoxat. Arzneim-Forsch/Drug Res 1983; 33:495–501.
99. Ivy GO, Schottler F, Wenzel J, Baudry M, Lynch G. Inhibitors of lysosomal enzymes: accumulation of lipofuscin-like dense bodies in the brain. Science 1984; 226:985–987.
100. Nandy K. Lipofuscinogenesis in mice early treated with centrophenoxine. Mech Ageing Dev 1978; 8:131–138.
101. Spoerri PE, Glees P. The effects of dimethylaminoethyl *p*-chlorophenoxyacetate on spinal ganglia neurons and satellite cells in culture. Mitochondrial changes in the aging neurons. An electron microscope study. Mech Ageing Dev 1974; 3:131–155.
102. Guili C, Bertoni-Freddari C, Pieri C. Morphometric studies on synapses of the cerebellar glomerulus: the effect of centrophenoxine treatment in old rats. Mech Ageing Dev 1980; 14:265–271.
103. Viticchi C, Gentile S, Piantanelli L. Centrophenoxine-

induced increase of beta-adrenoceptor density in brain cortex of old mice. Arch Gerontol Geriatr 1984; 3:77–82.

104. Nandy K. Centrophenoxine: effects on aging mammalian brain. J Am Geriatr Soc 1978; 26:74–81.

105. Hochschild R. Effect of dimethylaminoethyl *p*-chlorophenoxyacetate on the life span of male Swiss Webster albino mice. Exp Gerontol 1973; 8:177–183.

106. Gill PJ, Hanson DM. Effect of centrophenoxine on human embryonic lung cells in culture. In: Roberts J, Adelman RC, Cristofalo VJ, eds. Pharmacological intervention in the aging process. New York: Plenum Press, 1977:277–284.

107. Burgers PMJ. Eukaryotic DNA polymerases alpha and delta: conserved properties and interactions, from yeast to mammalian cells. Prog Nucleic Acid Res Mol Biol 1989; 37:235–280.

108. Busbee D, Sylvia V, Stec J, Cernosek Z, Norman J. Lability of DNA polymerase alpha correlated with decreased DNA synthesis and increased age in human cells. J Natl Cancer Inst 1987; 79:1231–1239.

109. Sylvia V, Curtin G, Norman J, Stec J, Busbee D. Activation of a low specific activity form of DNA polymerase alpha by inositol-1,4-bisphosphate. Cell 1988; 54:651–658.

110. Linn S, Kairis M, Holiday R. Decreased fidelity of DNA polymerase activity isolated from aging human fibroblasts. J Mol Biol 1981; 146:2818–2822.

111. Wang W, Warner HR, eds. Growth control during cell aging. Boca Raton, FL: CRC Press, 1989.

112. Frontiers in biology: the cell cycle. Science 1989; 246:603–640.

113. Gelfant S, Smith JG Jr. Aging: noncycling cells an explanation. Science 1972; 178:357–361.

114. Pendergrass WR, Saulewicz AC, Burmer GC, Rabinovitch PS, Norwood TH, Martin GM. Evidence that a critical threshold of DNA polymerase-alpha activity may be required for the initiation of DNA synthesis in mammalian cell heterokaryons. J Cell Physiol 1982; 113:141–151.

115. Lumpkin CK Jr, McClung JK, Pereira-Smith OM, Smith JR. Existence of high abundance antiproliferative mRNA's in senescent human diploid fibroblasts. Science 1986; 232:393–395.

116. Sugawara O, Oshimura M, Koi M, Annab LA, Barrett JC. Induction of cellular senescence in immortalized cells by human chromosome 1. Science 1990; 247:707–710.

117. Loo DT, Fuquay JI, Rawson CL, Barnes DW. Extended culture of mouse embryo cells without senescence: inhibition by serum. Science 1987; 236:200–202.

118. Maier JAM, Voulalas P, Roeder D, Maciag T. Extension of the life-span of human endothelial cells by an interleukin-1-alpha antisense oligomer. Science 1990; 249:1570–1574.

118a. West M. Reversing cellular aging in connective tissues. Life Extension Rep 1990; 10:73–75.

119. Lenaz G, ed. Coenzyme Q, biochemistry, bioenergetics, and clinical applications of ubiquinone. New York: Wiley, 1985.

120. Beyer RE, Burnett B-A, Cartwright KJ, et al. Tissue coenzyme Q (ubiquinone) and protein concentrations over the life span of the laboratory rat. Mech Ageing Dev 1985; 32:267–281.

121. Folkers K, Vadhanavikit S, Mortensen SA. Biochemical rationale and myocardial tissue data on the effective therapy of cardiomyopathy with coenzyme $Q_{10}$. Proc Natl Acad Sci USA 1985; 82:901–904.

122. Bliznakov EG. Coenzyme Q, the immune system and aging. In: Folkers K, Yamamura Y, eds. Biomedical and clinical aspects of coenzyme Q, vol. 3. New York: Elsevier/North-Holland, 1981:311–323.

123. Yamamura Y. A survey of the therapeutic uses of coenzyme Q. In: Lenaz G, ed. Coenzyme Q, biochemistry, bioenergetics, and clinical applications of ubiquinone. New York: Wiley, 1985:479–505.

124. Folkers K, Yamamura Y, eds. Biomedical and clinical aspects of coenzyme Q. New York: Elsevier/North-Holland, 1977.

125. Yamamura Y, Folkers K, Ito Y, eds. Biomedical and clinical aspects of coenzyme Q, Vol. 2. New York: Elsevier/North-Holland, 1980.

126. Folkers K, Yamamura Y, eds. Biomedical and clinical aspects of coenzyme Q, Vol. 3. New York: Elsevier/North-Holland, 1981.

127. Folkers K, Yamamura Y, eds. Biomedical and clinical aspects of coenzyme Q, Vol. 4. New York: Elsevier/North-Holland, 1984.

128. Folkers K, Yamamura Y, eds. Biomedical and clinical aspects of coenzyme Q, Vol. 5. New York: Elsevier/North-Holland, 1986.

129. Lenaz G, ed. Biochemistry, bioenergetics and clinical aspects of ubiquinone. London: Taylor & Francis, 1990.

130. Frei B, Kim MC, Ames BN. Ubiquinol-10 is an effective lipid-soluble antioxidant at physiological concentrations. Proc Natl Acad Sci USA 1990; 87:4879–4883.

131. Folkers K, Baker L, Richardson PC, et al. New progress on the biomedical and clinical research on coenzyme Q. In: K Folkers, Y Yamamura, eds. Biochemical and clinical aspects of coenzyme Q, Vol. 3. New York: Elsevier/North Holland, 1981:399–412.

132. Ozawa T. Formation of oxygen radicals in the electron transfer chain and antioxidant properties of coenzyme Q. In: Lenaz G, ed. Coenzyme Q, biochemistry, bioenergetics, and clinical applications of ubiquinone. New York: Wiley, 1985:441–456.

133. Demopoulos HB, Flamm ES. Possible pathology induced by $CoQ_{10}$. In: Folkers K, Yamamura Y, eds. Biomedical and clinical aspects of coenzyme Q, Vol. 3. New York: Elsevier North Holland, 1981:373–384.

134. Knoll J. Deprenyl (selegiline): the history of its development and pharmacological action. Acta Neurol Scand 1983; 95(Suppl):57–80.

135. Knoll J. Extension of life span of rats by long-term (−)-deprenyl treatment. Mt Sinai J Med 1988; 55:67–74.

136. Knoll J. The striatal dopamine dependency of life span in male rats. Longevity study with (−)deprenyl. Mech Ageing Dev 1988; 46:237–262.

137. Grimes JD, Hassan MN, Thakar J. Antioxidant therapy in Parkinson's disease. Can J Neurol Sci 1987; 14:483–487.

138. Finch CE. Age-related changes in brain catecholamines: a synopsis of findings in C57B1/6J mice and other rodent models. Adv Exp Biol Med 1978; 113:15–39.

139. Tetrud JW, Langston JW. The effect of deprenyl (Selegiline) on the natural history of Parkinson's disease. Science 1989; 245:519–522.

140. The Parkinson Study Group. Effect of deprenyl on the progression of disability in early Parkinson's disease. N Engl J Med 1989; 321:1364–1371.

141. Birkmayer W, Knoll J, Riederer P, Youdim MBH. (−)-Deprenyl leads to prolongation of L-dopa efficacy in

Parkinson's disease. Mod Probl Pharmacopsychiatry 1983; 19:170–176.

142. Birkmayer W, Knoll J, Riederer P, Youdim MBH, Hars V, Marton J. Increased life expectancy resulting from addition of L-deprenyl to Madopar treatment in Parkinson's disease: a longterm study. J Neural Transmission 1985; 64:113–127.

143. Sonntag WE, Forman LJ, Miki N, Trapp JM, Gottschall PE, Meites J. L-Dopa restores amplitude of growth hormone in old male rats to that observed in young male rats. Neuroendocrinology 1982; 34:163–168.

144. Chihara K, Kashio Y, Kita T, et al. L-Dopa stimulates release of hypothalamic growth hormone-releasing hormone in humans. J Clin Endocrinol Metab 1986; 62:466–473.

145. Boyd AE, Lebovitz HE, Pfeiffer JB. Stimulation of human-growth-hormone secretion by L-DOPA. N Engl J Med 1970; 283:1425–1429.

146. Cotzias GC, Miller ST, Nicholson AR Jr, et al. Prolongation of the lifespan in mice adapted to large amounts of L-Dopa. Proc Natl Acad Sci USA 1974; 71:2466–2469.

147. Cotzias GC, Miller ST, Tang LC, et al. Levodopa, fertility, and longevity. Science 1977; 196:549–550.

148. Tariot PN, Cohen RM, Sunderland T, et al. L-Deprenyl in Alzheimer's disease: preliminary evidence for behavioral change with monoamine oxidase B inhibition. Arch Gen Psychiatry 1989; 46:45–50.

149. Knoll J, Dallo J, Yen TT. Striatal dopamine, sexual activity and lifespan. Life Sci 1989; 45:525–531.

150. Milgram NW, Racine RJ, Nellis P, Mendonca A, Ivy GO. Maintenance on l-deprenyl prolongs life in aged male rats. Life Sci 1990; 47:415–420.

151. Youngster SK, McKeown KA, Jin Y-Z, Ramsay RR, Heikkila RE, Singer TP. Oxidation of analogs of 1-methyl-4-phenyl-1,2,3,6-tetrahydropyridine by monoamine oxidases A and B and the inhibition of monoamine oxidases by the oxidation products. J Neurochem 1989; 53:1837–1842.

152. Robinson DS. Changes in monoamine oxidase and monoamines with human development and aging. Fed Proc 1975; 34:103–107.

153. Landfield PW, Baskin RK, Pitler TA. Brain aging correlates: retardation by hormonal-pharmacological treatments. Science 1981; 214:581–584.

154. Gibbons A. Gerontology research comes of age. Science 1990; 250:622–625.

155. Baldwin E, Schultz PG. Generation of a catalytic antibody by site-directed mutagenesis. Science 1989; 245:1104–1107.

156. Breslow R. Artificial enzymes. Cold Spring Harbor Symp Quant Biol 1987; 52:75–81.

157. Eigler DM, Schweizer EK. Positioning single atoms with a scanning tunnelling microscope. Nature 1990; 344:524–526.

158. Drexler KE, Foster JS. Synthetic tips. Nature 1990; 343:600.

159. Utsugi Y. Nanometre-scale chemical modification using a scanning tunnelling microscope. Nature 1990; 347:747–749.

160. Jardetzky O, ed. Protein structure and engineering. New York: Plenum Press, 1989.

161. Narang SA, ed. Protein engineering. Approaches to the manipulation of protein folding. Boston: Butterworths, 1990.

162. Hahn KW, Wieslaw AK, Stewart JM. Design and synthesis of a peptide having chymotrypsin-like esterase activity. Science 1990; 248:1544–1547.

163. Drexler KE. Rod logic and thermal noise in the mechanical nanocomputer. In: Carter FL, Siatkowski R, Wohltjen H, eds. Molecular electronic devices. Amsterdam: Elsevier/North-Holland, 1988:39–56.

164. Drexler KE. Molecular machinery and molecular electronic devices. In: Carter F, ed. Molecular electronic devices II. New York: Marcel Dekker, 1988:549–571.

*Submitted for publication October 1989.*
*Final revision received February 1991.*

Author's addition in proof: Several recent publications have further amplified and extended many of the points discussed here. An exhaustive review of aging mechanisms by Finch (165) continues to implicate genetic factors in the genesis of senescence. Johnson recently showed that species-specific lifespan is genetically specified in *C. elegans* by isolating a mutant that ages at a much slower rate than wild-type populations (166). An intriguing synthesis has been proposed by Sohall and Allen (167, 168) that reinterprets evidence previously thought to support the free-radical hypothesis in favor of a genetic hypothesis of aging and suggests mechanisms for the genetic regulation of aging. Evidence that centrophenoxine improves protein synthesis in the spinal cord has been presented (169), and new attention has been given to the potential therapeutic utility of coenzyme $Q_{10}$ (170, 171).

Finally, the development of atomic switches (172) and insights into the utility of self-assembly for constructing molecular devices (173) further hasten nanotechnologic interventions.

### Acknowledgments

Supported in part by N.I.H. grants GM-1759 and BSRG 2507-RR 05737. The author thanks Alvin Steinberg for catalyzing the writing of this article.

165. Finch CE. Longevity, senescence, and the genome. Chicago: The University of Chicago Press, 1990.

166. Johnson TE. Increased life-span of age-1 mutants in *Caenorhabditis elegans* and lower Gompertz rate of aging. Science 1990; 249:908–912.

167. Sohall RS, Allen RG. Oxidative stress as a causal factor in differentiation and aging: a unifying hypothesis. Exp Gerontol 1990; 25:499–522.

168. Allen RG. Oxygen-reactive species and antioxidant responses during development: the metabolic paradox of cellular differentiation. Proc Soc Exp Biol Med 1991; 196:117–129.

169. Gupta SK, Sharma SP, James TJ, Reddy BN, Patro IK. Lipofuscinolytic activity of centrophenoxine: a possible explanation. AGE 1989; 12:152.

170. Greenberg SM, Frishman WH. Coenzyme $Q_{10}$: a new drug for myocardial ischemia. Med Clin North Am 1988; 72:243–258.

171. Greenberg SM, Frishman WH. Coenzyme $Q_{10}$: a new drug for cardiovascular disease. J Clin Pharmacol 1990; 30:596–608.

172. Eigler DM, Lutz CP, Rudge WE. An atomic switch realized with the scanning tunneling microscope. Nature 1991; 352:600–603.

173. Lindsey JS. Self-assembly in synthetic routes to molecular devices. Biological principles and chemical perspectives: a review. New J Chem 1991; 15:153–180.

*Student Corner*

# The Changing Controversy over Surgical Resection Margins for Stage I Cutaneous Melanoma

Michele S. Green, M.D.

Surgical excision is the standard mode of treatment of primary cutaneous melanomas. Since the first published report, in 1820, of a patient being cured by wide surgical excision (1), the traditional treatment of melanomas has been wide excision. Handley (2) is often credited with the fundamental principle of treating melanomas by excision with a 5-cm margin based on his studies of the distribution of tumor in the lymphatics surrounding cutaneous metastases.

Through the 1940s and until the late 1960s, malignant melanomas were being treated with resection margins of from 5 cm to 15 cm. The rationale for wide excision was based on two principles: the need to remove occult foci of malignant cells, and the fact that melanomas develop a field change involving an area of skin larger than the tumor. Wong and Cochran's studies (3, 4) of atypical melanocytes surrounding melanomas for 5-cm margins helped perpetuate the belief that wide excision was needed for cure.

In the past twenty years, controversy has continued over the resection margins needed for cure. The large disfiguring margins of the past have been found to be mostly unnecessary. Many variables affect prognosis for clinical stage I melanoma, including mitotic rate, Clark level, ulceration, and microscopic satellites. The most important variable in prognosis may be that of tumor "thickness." On the basis of thickness, tumors may be categorized as minimum, low, intermediate, or high risk. Breslow and Macht (5) were the first to show that in melanomas less than 0.76 mm thick, the rate of local recurrence was not influenced by the extent of resection margins. Many studies since have confirmed that tumors less than 1 mm thick need no more than 1–2 cm for an adequate resection margin. Mortality may be best related to the depth or thickness of the tumor. For tumors thicker than 2 mm, a resection margin greater than 3 cm will not influence prognosis.

The theories of melanoma growth that supported wide excision have been largely disproven. This review traces the history of and controversy over surgical resection margins for the treatment of malignant melanoma. In addition, current thinking on systemic dissemination, recurrence, microscopic satellites, and other important variables useful in estimating the prognosis of malignant melanoma is also reviewed.

## History of the Controversy

The first case of melanoma treated by surgical excision was published in 1820 by William Norris. In this first reported patient with nodular melanoma with satellite tumors, the tumor recurred in the scar in under six weeks (1). In 1857, Norris reported on eight additional cases of melanoma (6), including the first instance of a "cure" being attributed to wide surgical excision; the patient survived an additional eight years without recurrence of the disease. As a result of this case, Norris recommended surgery to "not only remove the disease, but cut away some of the healthy parts" (6). Additionally, he noted the helplessness of surgery when melanoma was already disseminated. Based on these cases, Norris was the first

---

Written when the author was a third-year medical student at the Mount Sinai School of Medicine (CUNY). Dr. Green is currently a first-year resident in Pediatrics at The New York Hospital—Cornell Medical Center. Address reprint requests to Michele S. Green, M.D., Dept. of Pediatrics, The New York Hospital, 525 E. 68th St., New York, NY 10021.

to record local recurrence of melanoma after excision, cite the need for surgical resection of a margin of normal healthy tissue, record epidemiologic information, and identify the futility of both medical and surgical treatment for disseminated melanoma.

Although Norris' case report highlights his astute powers of observation, he makes no estimate of what constitutes a reasonable margin of resection of normal tissue surrounding the primary lesion. Handley is credited with specifying a 5-cm area as the proper margin of resection. However, Handley's Hunterian lecture in 1907 (2) makes no such recommendations. Handley, in fact, recommends that the incision around a primary melanoma be made "about an inch from the edge of the tumor," in actuality 2.5 cm (2). Nevertheless, the standard margin continued to be 5 cm and wider.

Part of the explanation for the continued support of wide surgical excision lies in the historical context of the 1940s and 1950s, an atmosphere that encouraged wide surgical margins in surgical treatment of other cancers, such as Halsted's radical mastectomy for breast cancer (7). The wide acceptance of these large margins was also encouraged by the increased survival reported (8). Furthermore, there were those who recommended wide surgical excision because of microscopic evidence of residual tumor in the re-excision specimen (9). The fervor to increase the margins of resection was so great as to include surgical resection margins of up to 15 cm for stage III melanoma (10). Petersen (10), however, proposed only a 1-cm margin for stage I melanoma. Although Petersen attempted to have individual treatments determined by stage, he mistakenly assumed that this treatment consisted of offering wider surgical resection margins.

The arguments in favor of wide excision were also based on a misunderstanding of the histogenesis of malignant melanoma. Again, Handley can be credited for first attempting to understand the nature of melanoma metastasis. He concluded that "permeation of the lymphatics is the principal agent in this local centrifugal spread and that it occurs primarily and most extensively in the plane of the deep fascial lymphatic plexus" (2). Since Handley studied melanoma only within the context of metastases, and not on the basis of a primary lesion, it is no surprise that he witnessed spread along the deep fascia plane. These misunderstandings of melanoma spread continued. For example, Petersen's notion of wide surgical excision was based on his concept of melanoma growth: "The pigmented flare or halo in which the tumor appears is all malignant and seems to spread by the communication of malignant potential to neighboring epidermal melanocytes" (10). Petersen based his recommendation of a 15-cm resection margin on this reasoning.

Two of the most influential papers that argued for aggressive surgical resection were those of Wong (3) and Cochran (4). They studied the distribution of melanocytes in the normal skin surrounding melanomas. Cochran proposed the notion of a "field change" around a melanoma and noted a high incidence of melanocytes in the skin surrounding many melanomas. Based on the presence of these atypical melanocytes, he justified the use of wide margins of excision (4). Wong demonstrated atypical melanocytes in the clinically normal skin within a 5-cm radius from the edge of the melanomas (3). Wong reasoned that melanocytes communicated with each other through their dendritic processes. He believed that the atypical melanocytes represented melanocytes whose changes had been "induced" by the primary melanoma. Thus, based on misinterpreting the nature of these atypical melanocytes, he argued for a 5-cm resection margin.

Our current understanding of atypical melanocytes is due largely to Ackerman's studies on these topics. Atypical melanocytes can be found in many benign conditions, and their presence does not automatically confer malignancy (11). There are many examples of melanocytic lesions that possess atypical melanocytes and unusual patterns of melanocytes within the epidermis, such as simple lentigo, Spitz's nevus (junctional type), and congenital melanocytic nevi (11, 12). Furthermore, the use of immunohistochemical markers, such as HMB-45 in the diagnosis of melanoma, has differentiated between lesions possessing atypical melanocytes and the lesions of melanoma (13, 14).

Thus, the prejudice over the past century has been to practice wide surgical excision. The current recommendations for excision by the Massachusetts General Hospital for stage I melanomas are based on tumor thickness (15):

| Tumor thickness | Excision margin |
|---|---|
| <1.0 mm | 1 cm |
| 1.0–1.5 mm | 1.5 cm |
| 1.5–4.0 mm | 3 cm |
| >4.0 mm | 3 cm |

Over the last two decades the decreases in surgical resection margins can be largely attributed to the work of Breslow (5) and others.

## Variables Affecting Prognosis

### Tumor Thickness

Olsen was the first to conclude that survival and local recurrence were not affected by the degree of surgical resection (16). However, Olsen's work did not correct for the thickness of the tumor and the possibility that 1-cm margins were taken for thinner tumors, whereas 5-cm margins were used for the larger and thicker melanomas. Breslow was the first to identify the significance of tumor thickness as the basis for prognosis (17).

In Breslow's study of resection margins for thin melanomas, he concluded that the rate of local recurrence of melanomas thinner than 0.76 mm was not influenced by the extent of resection margin (5). He helped dispel the histopathologic theories of Wong and Cochran by refuting the "field change" theory and the significance of atypical melanocytes surrounding malignant melanomas: "If these are malignant cells, a 5 centimeter margin should not be adequate for these tumors, and many of our patients with thin melanomas should have had local recurrences" (5). Breslow highlighted the inconsistency in assigning to all atypical melanocytes malignant status and in basing surgical resection margins on their presence.

In order to clearly determine what effect, if any, resection margins have on prognosis, four separate points must be addressed: the influence of resection margins on local recurrence, on local regional lymph node metastases, on distant metastases, and on survival.

**Local Recurrence.** Balch confirmed the fact that in tumors less than 0.76 mm thick, the margins of resection had no influence on the incidence of local recurrence (18). Bagley found the local recurrence rate of these thin tumors to be 3% (19), similar to Breslow's rate of 0% (17). Numerous articles cite these same low values for local recurrence of thin tumors irrespective of the size of the margins used: Urist, 0.1% (20); Cascinelli, 1% (21); Elder, 0% (22); Roses, 0% (23). From these studies, the margin of resection for tumors less than 1 mm thick has been proposed at 1 cm (14).

Controversy still surrounds local recurrence for tumors thicker than 1 mm. Cascinelli identified an "association" between resection margins and recurrences; the percentage of recurrences decreased from 10.8% to 3.0% as resection margins increased from 1 cm to 5 cm (21). Cascinelli interpreted this trend as due to differing ratios of thick to thin tumors within the various groups of resection margins. He concluded that for tumors less than 2.0 mm thick, the frequency of local recurrence was unrelated to resection margins but cautioned against the use of narrow margins for lesions greater than 2 mm thick (21). A later study confirmed that narrow excision was adequate for thin lesions but thicker tumors had a local recurrence rate 2.5 times higher with narrow excision (24).

**Metastases.** The second question concerned the effect of surgical resection margins on metastases. Breslow was the first to emphasize that metastases were not influenced by the extent of resection, but rather directly proportional to tumor thickness (5). Balch confirmed Breslow's findings and noted that the incidence of metastases, both to regional lymph nodes and to distant sites, was directly proportional to tumor thickness (18). Schmoekel proved that the incidence of metastases was independent of the margins of resection (25). He found that most cases of metastases involved patients who were already at a high risk for developing metastases based on increased tumor thickness. Veronesi concurred with these findings and offered evidence of no difference in metastatic disease involving regional nodes and distant organs in those treated with narrow excision (26).

**Survival.** With respect to the issue of survival, Hansen and McCarten (27) showed no difference in survival in patients with melanoma less than 1.5 mm thick who had simple excision and those who had wide excision. Cascinelli (21) reported that survival was not influenced by the size of the resection margin, but rather was best predicted by tumor thickness. Many studies (23, 24, 27, 28) have duplicated these findings of the World Health Organization Study. Only one study, by Aitken, contends that increasing the surgical resection margin to 2.5 cm improved survival, but found that no further protection was afforded by margins over 3 cm (29).

**"Cryptic" Dissemination.** In summary, these studies imply that when a tumor can no longer be considered thin, it is more apt to metastasize regardless of the extent of surgical resection. If large excisions could reduce the incidence of distant metastases, overall survival would improve (30). It would thus appear that a very thick lesion may already portend this "cryptic" dissemination, and wider excision will be of little or no benefit. Identification of tumor thickness and of any accompanying variables and markers that can portend this "cryptic" dissemination will prevent the need for fruitless surgical resection.

### Other Variables

Other variables in addition to thickness have been suggested as portending a poorer prognosis,

TABLE

*Changes in Recommended Surgical Resection Margins for Malignant Melanoma, 1857–1990*

| Year | Author (ref) | Recommended surgical resection margin |
|---|---|---|
| 1857 | Norris (6) | Margin of normal healthy tissue, unspecified size |
| 1907 | Handley (2) | 2.5 cm |
| 1952 | Pack (49) | Wide local excision, unspecified size |
| 1955 | Lund (8) | Wide local excision, unspecified size |
| 1962 | Petersen (10) | 1 cm: Stage I<br>5–7 cm and deep fascia: Stage II<br>6–15 cm and deep fascia: Stage III |
| 1966 | Olsen (16) | 1–5 cm |
| 1970 | Wong (3) | 5 cm |
| 1971 | Cochran (4) | Wide local excision, unspecified size |
| 1977 | Breslow (5) | <2 cm for lesions <0.76 mm |
| 1979 | Balch (18) | Wide local excision, 3–5 cm |
| 1982 | Day (50) | 1.5 cm for lesions <0.85 mm & non-BANS lesions 0.85–1.69 mm |
| 1983 | Aitken (29) | 2 cm for lesions <0.76 mm;<br>3 cm for all thicker lesions |
| 1983 | Schmoeckel (25) | 2 cm for lesions <0.76 mm;<br>3 cm for all thicker lesions |
| 1983 | Elder (22) | 1 cm for low-risk lesions (level II, III, lesions <0.76 mm, lesions without regression)<br>3–5 cm for intermediate-risk lesions (0.76–1.5 mm)<br>3–5 cm for high-risk lesions (all lesions >1.5 mm) |
| 1984 | Kelly (45) | 1 cm for lesions <1.0 mm<br>3 cm for lesions >1.0 mm |
| 1985 | Urist (22) | 1–2 cm for lesions <1 mm<br>2–3 cm for lesions >1 mm |
| 1988 | Veronesi (26) | 1 cm for lesions <2 mm |
| 1990 | Ho (15) | 1 cm for lesions <1.0 mm<br>1.5 cm for lesions 1.0–1.5 mm<br>3 cm for lesions >1.5 mm |

including mitotic rate, age, sex, anatomic site, ulceration, and microscopic satellites. Some point to increasing age as an indicator of poorer prognosis (31–34), yet, after correlation for thickness, age has not been shown to be such an indicator (25, 35–37). Additionally, some studies show better survival rates for women (34). The presence of mitoses has been shown to indicate a poorer prognosis in melanomas of intermediate thickness (36) and in thick melanomas greater than 4.0 mm thick (38). Survival seems to decrease with the presence of these mitoses, the potential for metastases being attributed to tumors with very high mitotic rates (greater than 25 mitoses/mm$^2$) (39). Ulceration has been shown to be an adverse prognostic sign, with a decrease in five-year survival from 80% to 50% (40).

**Microscopic Satellites and "Cryptic" Dissemination.** The significance of microscopic satellites remains the most controversial and elusive of these variables. The presence of microscopic satellites has often been cited as a reason for wide surgical excisions, especially for tumors greater than 2.25 mm in thickness (22). Hence, it is important to understand microscopic satellites from both a histologic and a therapeutic outlook. Microscopic satellites have been defined as discontinuous tumor deposits that are within 5 cm of the primary lesion at the time of surgery (22), yet this definition underlies the controversy as to the nature of these microscopic satellites: "Conceptually it is uncertain whether microscopic satellites represent serpentine tongues of tumor originating from the depths of biologically aggressive melanomas, 'in transit' dermal metastases, or vascular plugs that have obliterated their containing structure" (41). Although there is histologic disagreement, it is generally recognized that these satellites can be classified as elements of the vertical growth phase (22, 41).

The currently recognized pathobiology of malignant melanoma includes both a radial growth phase (RGP) and a vertical growth phase (VGP), tumors that display only a RGP pattern having no chance of metastasis (42–44). As outlined above, other histologic parameters of VGP, such as ulceration and mitoses, all have adverse prognostic significance (45). Although some studies claim that the presence of microscopic satellites decreases survival (44, 46), others show no effect on five-year survival (22, 47). It has been shown that increasing tumor thickness is accompanied by increasing numbers of microscopic satellites, and lesions less than 2 mm thick have few microscopic satellites (48). Harrist (47) found that primary melanomas greater than 1.50 mm thick with microscopic satellites were more often associated with occult regional lymph node metastases than were melanomas of similar thickness without satellites—53% versus 12% (47). Thus, the vital question that remains is not simply whether the presence of microscopic satellites decreases survival, but whether their presence is a marker for "cryptic" dissemination.

## Summary

Over the last century, the recommended margins of surgical resection for cutaneous malignant melanoma have decreased (Table). There is a danger in misinterpreting histopathologic findings such as atypical melanocytes and microscopic satellites as the basis for choosing wide surgical excision. The efficacy of wide surgical excision for malignant melanoma remains to be proven. When tumor thickness is disregarded, metastasis and overall survival seem unaffected by the width of surgical margins.

The surgical treatment of malignant melanoma should be complete excision of the neoplasm. If the margins are clear histopathologically, no further local surgery is needed. Wide surgical resection of the primary lesion will have no effect on distant metastases (30). Once clear markers of melanoma dissemination have been identified, based either on thickness or on other histopathologic or immunologic variables, an adequate resection margin can be formulated. A thorough understanding of exactly which variables portend "cryptic" dissemination and influence prognosis will lead to better management of cutaneous malignant melanoma.

## Acknowledgments

Special thanks to Dr. Robert Phelps for his kind support.

## References

1. Norris W. Case of a fungoid disease. Edinburgh Med Surg J 1820; 16:562.
2. Handley WS. The pathology of melanocytic growth in relation to their operative treatment: lecture 1. Lancet 1907; 1:927–933, 996–1003.
3. Wong CK. A study of melanocytes in the normal skin surrounding malignant melanoma. Dermatologica 1970; 141:215–225.
4. Cochran AJ. Studies of the melanocytes of the epidermis adjacent to tumors. J Invest Dermatol 1971; 57:38–43.
5. Breslow A, Macht SD. Optimal size of resection margins for thin cutaneous melanomas. Surg Gynecol Obstet 1977; 145:691–692.
6. Norris W. Eight cases of melanosis with pathological and therapeutic remarks on that disease. London: Longman, Brown, Longmans, and Roberts, 1857.
7. Zitelli JA, Mohs FE, Larson P, et al. Mohs micrographic surgery for melanoma. Dermatol Clin 1980; 7:833–843.
8. Lund RH, and Inhen M. Malignant melanoma: clinical and pathological analysis of 93 cases. Surgery 1955; 38: 652–659.
9. Mundth ED, Guralnick EA, Raker JW. Malignant melanoma: clinical study of 427 cases. Ann Surg 1965; 162:15–28.
10. Petersen NC, Bodenham DC, Lloyd OC. Malignant melanomas of the skin: a study of the origin, development, etiology, spread, treatment, and prognosis. Br J Plast Surg 1962; 15:49–116.
11. Kamino H, Ackerman AB. The problem of interpreting the meanings of atypical melanocytes and unusual patterns of melanocytes with the epidermis. In: Ackerman AB, ed. Pathology of malignant melanoma. New York: Masson Publishing, 1981:129–157.
12. Ackerman AB. Melanocytic proliferations that simulate malignant melanoma histopathologically. In: Mihm MC Jr, Murphy GF, Kaufman N, eds. Pathobiology and recognition of malignant melanoma. Baltimore: Williams and Wilkins, 1988:153–173.
13. Wick M, Swanson P, Rocamora A. Recognition of malignant melanoma by monoclonal antibody HMB-45: An immunohistochemical study of 200 paraffin-embedded cutaneous tumors. J Cutan Pathol 1988; 15:201–207.
14. Ordonez N, Ji X, Hickey R. Comparison of HMB-45 monoclonal antibody and S-100 protein in the immunohistochemical diagnosis of melanoma. AM J Clin Pathol 1988; 90:385–390.
15. Ho VC, Sober AJ. Therapy for cutaneous melanoma: an update. J Am Acad Dermatol 1990; 22:159–176.
16. Olsen G. The malignant melanoma of the skin-new theories based on a study of 500 cases. Acta Chir Scand 1966; 365(Suppl):1–220.
17. Breslow A. Thickness, cross-sectional areas and depth of invasion in the prognosis of cutaneous melanoma. Ann Surg 1970; 172:902–908.
18. Balch CM, Murad TM, Soong S, et al. Tumor thickness as a guide to surgical management of clinical stage I melanoma patients. Cancer 1979; 43:883–888.
19. Bagley FH, Cady B, Lee A, et al. Changes in clinical presentation and management of malignant melanoma. Cancer 1981; 47:2126–2134.
20. Urist MM, Balch CM, Soong, S, et al. The influence of surgical margins and prognostic factors predicting the risk of local recurrence in 3445 patients with primary cutaneous melanoma. Cancer 1985; 55:1398–1402.
21. Cascinelli N, van der Esch EP, Breslow A, et al. Stage I melanoma of the skin: the problem of resection margins. Eur J Cancer 1980; 16:1079–1085.
22. Elder DE, Guerry D IV, Heiberger RM, et al. Optimal resection margin for cutaneous malignant melanoma. Plast Reconstr Surg 1983; 71:66–72.
23. Roses DF, Harris MN, Rigel D, et al. Local and in-transit metastases following definitive excision for primary cutaneous malignant melanoma. Ann Surg 1983; 198:65–69.
24. Milton GW, Shaw HM, McCarthy WH. Resection margins for melanoma. Aust NZ J Surg 1985; 55:225–226.
25. Schmoekel C, Bockelbrink A, Bockelbrink H, et al. Low- and high-risk malignant melanoma—III: prognostic significance of the resection margin. Eur J Cancer Clin Oncol 1983; 19:245–249.
26. Veronesi U, Cascinelli N, Adamus J, et al. Thin stage I primary cutaneous malignant melanoma: comparison of excision with margins of 1 or 3 cm. N Engl J Med 1988; 318:1159–1162.
27. Hansen MG, McCarten AB. Tumor thickness and lymphocytic infiltration in malignant melanoma of the head and neck. Am J Surg 1974; 128:557–561.
28. Wanebo HJ, Woodruff J, Fortner JG. Malignant melanoma of the extremities: a clinicopathologic study using levels of invasion (microstage). Cancer 1975; 35:666–676.
29. Aitken DR, Clausen K, Klein JP, et al. The extent of primary melanoma excision: a reevaluation—how wide is wide? Ann Surg 1983; 198:634–641.
30. Ackerman AB, Scheiner AM. How wide and deep is wide and deep enough? Hum Pathol 1983; 14:743–744.
31. Sober AJ, Day CL, Fitzpatrick TB, et al. Early death from clinical stage I melanoma. J Invest Dermatol 1983; 80:50s–52s.
32. McCarthy WH, Shaw HM, Milton GW. Melanoma in New South Wales, Australia: experience at the Sidney Mel-

anoma Unit. In: Balch CM, Milton GW, Shaw HM, Soong S, eds. Cutaneous melanoma: clinical management and treatment results worldwide. Philadelphia: Lippincott, 1985:371–387.

33. Balch CM, Soong S, Shaw HM, et al. An analysis of prognostic factors in 4000 patients with cutaneous melanoma. In: Balch CM, Milton GW, Shaw HM, Soong S, eds. Cutaneous melanoma: clinical management and treatment results worldwide. Philadelphia: Lippincott, 1985:321–352.
34. Thorn M, Adami HO, Ringborg U, et al. Long term survival in malignant melanoma with special reference to age and sex as prognostic factors. J Natl Cancer Inst 1987; 79:969–974.
35. Day CL Jr, Mihm MC Jr, Sober AJ, et al. Prognostic factors for melanoma patients with lesions 0.76–1.69 mm in thickness: An appraisal of "thin" level IV lesions. Ann Surg 1982; 195:30–34.
36. Day CL Jr, Mihm MC Jr, Lew RA, et al. Prognostic factors for patients with clinical stage I melanoma of intermediate thickness (1.51–3.99 mm). Ann Surg 1982; 195:35–43.
37. Day CL Jr, Lew RA, Mihm MC, et al. A multivariate analysis of prognostic factors for melanoma patients with lesions ≥3.65 mm in thickness. Ann Surg 1982; 195:44–49.
38. van der Esch EP, Cascinelli N, Preda F, et al. Stage I melanoma of the skin: evaluation of prognosis according to histologic characteristics. Cancer 1981; 48:1668–1673.
39. Schmoekel C, Bockelbrink A, Bockelbrink H, et al. Low and high risk melanoma: 1. Evaluation of clinical and histologic prognosticators in 585 cases. Eur J Cancer Clin Oncol 1983; 19:227–235.
40. Balch CM, Wilkerson JA, Murad TM, et al. The prognostic significance of ulceration of cutaneous melanoma. Cancer 1980; 45:3012–3017.
41. Murphy GF, Mihm MC Jr. Histologic reporting of malignant melanoma. In: Mihm MC Jr, Murphy GF, Kaufman N, eds. Pathobiology and recognition of malignant melanoma. Baltimore: Williams and Wilkins, 1988:79–93.
42. Clark WH, Ainsworth AM, et al. The development biology of primary human malignant melanomas. Semin Oncol 1975; 2:83–103.
43. Clark WH, From L, Bernardino EA. The histogenesis and biological behavior of primary human malignant melanoma of the skin. Cancer Res 1969; 29:705–727.
44. Imber MJ, Mihm MC Jr. Biological and prognostic significance of vertical growth phase characteristics in malignant melanoma. In Mihm MC Jr, Murphy GF, Kaufman N, eds. Pathobiology and recognition of malignant melanoma. Baltimore: Williams and Wilkins, 1988:19–34.
45. Kelly JW, Sagebiel RW, Calderon W, et al. The frequency of local recurrence and microsatellites as a guide to reexcision margins for cutaneous malignant melanoma. Ann Surg 1984; 200:759–763.
46. Day CL, Harrist TJ, Gorstein F, et al. Malignant melanoma: prognostic significance of "microscopic satellites" in the reticular dermis and subcutaneous fat. Ann Surg 1981; 194:108–112.
47. Harrist TJ, Rigel DS, Day CL. Microscopic satellites are more highly associated with regional lymph node metastases than is primary melanoma thickness. Cancer 1984; 53:2183–2187.
48. Shaw JHF. Malignant melanoma in Aukland, New Zealand. Surg Gynecol Obstet 1988; 166:425–430.
49. Pack GT, Gerber DM, Scharnagel IM. End results in the treatment of malignant melanoma: a report of 1190 cases. Ann Surg 1952; 136:905–911.
50. Day CL, Mihm MC, Sober AJ, et al. Narrower margins for clinical stage I malignant melanoma. N Engl J Med 1982; 306:479–482.

*Submitted for publication March 1991.*

# Abstracts

*The abstracts presented here, all by principal investigators in the Department of Medicine of the Mount Sinai School of Medicine (which includes the affiliates Beth Israel Medical Center, the Bronx Veterans Administration Medical Center, and Elmhurst Hospital Center), in New York City will be exhibited as posters in the Department of Medicine Ninth Annual Research Day, October 22, 1991. Only those abstracts by principal investigators who have consented to publication in the* Journal *are included; others are listed and referenced.*

*Six abstracts by students, fellows, and house staff working in the Department of Medicine have received Ninth Annual Department of Medicine Research Awards and will be presented orally at a Special Grand Rounds on October 22; these constitute the last group of abstracts published here.—Robert Safirstein*

# Research Day in Medicine 1991

## Alcohol Research and Treatment Center

1. **Lack of an Alcohol Dehydrogenase (ADH) in the Stomach of Japanese Subjects.** E Baraona, T Takagi,* L Shoichet, A Yokoyama,* H Ishii,* CS Lieber, Section of Liver Disease & Nutrition, Alcohol Research & Treatment Ctr., 130 West Kingsbridge Road, Bronx VA Medical Ctr, Bronx, NY 10468, (212) 579-1646; Mount Sinai School of Medicine (CUNY), NY 10029, and *Keio University, 35 Shinanomachi, Shinjuku-ku, Tokyo 160, Japan.

We recently reported that the human stomach plays a role in ethanol metabolism and contains 3 classes of ADH isozymes.

To investigate racial differences, we assessed gastric ADH isozymes in 4 white and 4 black male Americans and 13 male Japanese patients. The nonoriental group consisted of 6 patients operated on for morbid obesity, 1 for gastric ulcer, and 1 for cancer. The oriental samples consisted of 9 cases operated on for gastric cancer, 1 for ulcer, and 3 autopsies (with terminal liver disease).

Regardless of the underlying condition, no difference was found between white and black Americans. In addition to low Km Class I isozymes and high Km Class III isozymes (similar to those present in human liver), there was another isozyme class with high Km for ethanol and slower cathodic mobility than Class I isoenzymes on starch gel electrophoresis. By DEAE-Sephadex chromatography, it was eluted with the equilibration buffer after Class I isozymes. It was not present in the human liver and seemed to correspond to a gastric isozyme recently purified, characterized, and defined as $\sigma$ADH by Moreno & Parés (1991), which resembles but is not identical to Class-II or $\pi$ADH. By contrast, 9 of the 13 Japanese stomachs lacked this isozyme and had only those of Class I and III. This lack of isozyme was also evident on DEAE-sephadex: only a single peak of activity (reacting more strongly with 50 mM than with 2 M ethanol) was eluted in the buffer wash immediately after the void volume. We do not as yet know whether the absence of this isozyme is due to differences in dietary habits of Japanese and Americans or represents a genetic trait, which may be related to differences in gastric pathophysiology.

In conclusion, an ADH isozyme, which is prominent in the gastric mucosa of nonoriental subjects, is absent in most of the Japanese stomachs studied. This finding may eventually lead to elucidation of the physiologic role of this isozyme.

2. **Isoelectric Focusing/Western Blotting: A Novel and Practical Method for Quantitation of Carbohydrate-Deficient Transferrin in Alcoholics.** Y Xin, JM Lasker, CS Lieber, Section of Liver Disease & Nutrition, Alcohol Research & Treatment Ctr., 130 West Kingsbridge Road, Bronx VA Medical Ctr, Bronx, NY 10468 (212) 579-1646, Mount Sinai School of Medicine (CUNY), NY 10029. Alcoholism: Clin Exp Res 1991; 15:376.

3. **The Effects of Aspirin, Acetaminophen, and Ibuprofen on Blood Alcohol Concentrations, Alcohol Absorption, and Gastric ADH Activity.** RT Gentry, R Roine, R Lim Jr, E Baraona, CS Lieber, Section of Liver Disease & Nutrition, Alcohol Research & Treatment Ctr., 130 West Kingsbridge Road, Bronx VA Medical Ctr, Bronx, NY 10468, (212) 579-1646, Mount Sinai School of Medicine (CUNY), NY 10029.

Previous findings demonstrated that 1 g of aspirin given with a standard breakfast increases peak blood alcohol concentrations (BAC) to 139% of control, and the area under curve (AUC) to 126% of control following a 0.3 g/Kg dose of alcohol consumed one hour later. The present study was carried out to determine the mechanisms involved, and to compare the results with the effects of other common analgesics.

Pharmacokinetic analysis of the BAC time curve indicates that aspirin induced a statistically significant increase in the bioavailability of alcohol, consistent with its inhibition of gastric alcohol dehydrogenase (ADH) activity in vitro. This analysis also demonstrated an increase in the apparent rate of alcohol absorption ($p < 0.05$), suggestive of an additional mechanism. Acetaminophen clearly inhibited ADH activity of human gastric mucosa (to 43% of baseline at 5 mM and to 33% of baseline at 10 mM, using 500 mM ethanol, $p < 0.05$). However, 1 g of acetaminophen given to 5 men did not significantly increase blood alcohol concentrations as measured by either the peak BAC (113% of control) or the AUC (119% of control, n.s.). In contrast to both aspirin and acetaminophen, ibuprofen (up to 10 mM) had no inhibitory effect on gastric ADH activity. Five men given ibuprofen (400 mg) showed nonsignificant increases of peak BAC (110% of control, n.s.) and AUC (111% of control, n.s.). Pharmacokinetic analysis indicated that neither ibuprofen nor acetaminophen significantly increased the apparent rate of alcohol absorption nor the bioavailability of alcohol.

These findings demonstrate that aspirin increases blood alcohol concentrations more than either of the other two analgesics. Furthermore, there appear to be two mechanisms by which aspirin affects peak alcohol concentrations and AUC, one which involves inhibition of gastric ADH activity, and one which increases the rate of alcohol absorption.

## Cardiology

4. **Should Transesophageal Echocardiography Be Performed on All Stroke Patients?** JA Stahl, EA Fisher, RL Taikowski, JH Budd, AS Klapper, S Turhim, D Horowitz, ME Goldman, Division of Cardiology,

Department of Medicine, Box 1030, Mount Sinai Medical Center, One Gustave L. Levy Place, New York, NY 10029, (212) 241-3078.

Cardiac embolic sources of cerebrovascular accidents (CVA) may originate from the ventricles, valves, atria/appendages, aorta, or intracardiac shunts. Transesophageal echo (TEE) can resolve details of cardiac chambers not imaged by transthoracic echo (TTE).

Thus, we compared 200 consecutive patients who underwent both TTE and TEE: 76 with definite stroke and 124 with other diagnoses. Strokes were CAT-scan-documented 1 to 7 days prior to TEE.

Findings by the 2 methods were compared. Echocardiogram results were categorized into either high risk for stroke (atrial septal defect [ASD], patent foramen ovale [PFO], thrombus, vegetation, mass) or intermediate risk (left atrial [LA] "smoke," ascending aortic plaque [AoP], atrial septal aneurysm [ASA]). (* = $p < .02$ ** = $p < .001$)

| | # | Afib | PFO | LA-smoke | ASA | AoP |
|---|---|---|---|---|---|---|
| CVA+ | 76 | 12 | 7 | 14 | 5 | 22 |
| | | | | * | | ** |
| CVA− | 124 | 33 | 2 | 18 | 0 | 31 |

Of the 76 stroke patients, 34 (45%) had high-risk echo findings, and 13 (17%) intermediate risk. Importantly, TEE provided unique information not seen by TTE in 52 of 76 (68%) stroke patients.

Therefore, TEE is invaluable in the cardiac evaluation of stroke patients.

**5. Left Ventricular Mass Is Related to Insulin Resistance in Mild Hypertension.** RA Phillips, S Zimmerman, DT Finegood,* S Shimabukuro,† M Ardeljan, C Wilkenfeld, A Dunaif, R Gorlin, LR Krakoff, Divisions of Hypertension, Cardiology and Endocrinology, Box 1085, Mount Sinai School of Medicine, NY, NY 10028, (212) 241-4691; *University of Alberta, Edmonton, Canada; †Hirano Hospital, Gifu, Japan.

In mild hypertension, factors other than arterial pressure also contribute to left ventricular (LV) mass. The separate contributions to LV mass index (LVMI, $g/m^2$) of daily arterial pressure and insulin resistance were determined in nonobese, nondiabetic mild hypertensives.

Eighteen untreated subjects (body mass index [BMI] $\leq 28$ $kg/m^2$) were studied by 24-hr ambulatory blood pressure monitoring (systolic [SBP], diastolic [DBP] mm Hg) and 2-D directed m-mode echocardiography. Insulin sensitivity index ([SI] $\times 10^{-4}$ $min^{-1}$ [$\mu U/mL$]) was determined from the frequently sampled IV glucose tolerance test. Characteristics (mean ± SE) for the group are:

| AGE | BMI | SI | SBP | DBP | LVMI |
|---|---|---|---|---|---|
| 43 ± 3 | 25 ± .5 | 4.0 ± .5 | 133 ± 3 | 85 ± 2 | 87 ± 4 |

LVMI was significantly related to SBP ($r = 0.50$, $p < 0.05$) and DBP ($r = 0.56$, $p < 0.05$). LVMI was significantly related to SI ($r = -0.55$, $p < 0.05$). In multivariate stepwise analysis, after adjustment for BP, BMI, and age, a significant relationship remained between LVMI and SI ($p = 0.02$). In this model, for each unit decrease in SI, LVMI increases by 4.2 $g/m^2$.

Thus, in nonobese mild hypertension, insulin sensitivity influences LV mass independently of BP. Prevention of LV hypertrophy and optimal regression of LV mass may require reduction of BP and improved insulin sensitivity.

**6. A Reassessment of the Prognostic Implications of Vegetation Size and Prolonged Fever in Right-Sided Endocarditis in Intravenous Drug Abusers.** SR Hecht, M Berger, Division of Cardiology, Department of Medicine, Beth Israel Medical Center, First Avenue at 16 Street, New York, NY 10003, (212) 420-2092.

Controversy still exists concerning the prognostic implications of vegetation size >1.0 cm and prolonged fever >3 weeks in right-sided endocarditis.

To further define the influence of these factors on in-hospital mortality, we reviewed the records and echocardiograms of all IV drug abusers who were hospitalized between 1978 and 1986 with clinical right-sided endocarditis and vegetations detected on native right-sided valves.

There were 98 episodes in 92 pts (tricuspid, 93; pulmonic, 4; both valves, 1). Vegetation size ranged from 0.5–4.3 cm (mean 1.5 ± 0.7 cm). Vegetations >1.0 cm occurred in 81% and fever >3 wks in 24%. Overall mortality was 7% and no patient underwent surgery. Relationships between vegetation size, prolonged fever, and mortality were:

| | VEG size (cm) | | | fever (wk) | |
|---|---|---|---|---|---|
| | <1.0 | 1.1–2.0 | >2.0 | <3 | >3 |
| deaths/ # pts | 0/19 | 1/61 | 6/18 | 5/74 | 2/24 |
| % mortality | 0 | 2 | 33* | 7 | 8** |

* $P = 0.0002$, **$P$ = NS

*Conclusions:* Native right-sided endocarditis has an overall favorable prognosis. The presence of vegetation size >1.0 cm and fever >3 weeks are common in right-sided endocarditis and are not predictive of a poor outcome. Vetegations >2.0 cm are associated with increased in-hospital mortality.

**7. The Importance of the Contractile Function of the Left Atrial Appendage.** ME Goldman, JH Budd, R Taikowski, EA Fisher, Division of Cardiology, Department of Medicine, Box 1030, Mount Sinai School of Medicine, New York, NY 10029, (212) 241-3078.

The left atrial appendage (LAA) is considered an accessory, passive structure, important only in its potential for thrombus formation in patients with atrial fibrillation (AFIB).

To evaluate the function of the LAA, we performed thransthoracic and transesophageal (TEE) with pulsed and color Doppler in 38 patients who had no significant mitral regurgitation. We measured LA and LAA area ($cm^2$), peak LAA flow (PK, cm/sec), LAA velocity time integral (VTI), and heart rate (HR); NSR = normal sinus rhythm, * = $p < .0001$, ** = $p < .003$:

| | # | HR | LA | LAA | PK | VTI |
|---|---|---|---|---|---|---|
| NSR | 20 | 73 | 20.2 | 4.4 | 54.4 | 3.3 |
| | | | ** | * | * | * |
| AFIB | 18 | 83 | 28.2 | 7.5 | 22.8 | 1.5 |

LAA of AFIB patients had rapid (>400/min) spikes with significantly lower velocity. Also LA "smoke" was seen in 14/18 AFIB patients and only 2/20 NSR patients, with more stagnant smoke in the LAA. Active LAA contraction produced forward, not reflected flow, documented visually and by color and pulsed Doppler flow even in AFIB pts.

Thus, patients with AFIB may be at greater risk for LAA thrombus formation due to stasis of blood in an enlarged and severely hypocontractile left atrial appendage.

**8. Left Atrial "Smoke" Is a Major Risk Factor for Thrombus Formation.** EA Fisher, JA Stahl, JH Budd, RA Goldman, A Klapper, Y Lee, ME Goldman, Division of Cardiology, Department of Medicine, Box 1030, Mount Sinai School of Medicine, 1 Gustave L. Levy Place, New York, NY 10029, (212) 241-3078.

The presence of "smoke" in dilated (↑) cardiac chambers, particularly the left atrium and appendage (LA/A), is frequently associated with low flow states and possibly thrombus (THR) formation. Transesophageal echo (TEE), imaging at a higher frequency and with less attenuation than transthoracic echo (TTE), may be superior for smoke and THR detection in the LA/A.

To determine the incidence of smoke and its relationship to THR, we performed TEE immediately after standard TTE in 410 consecutive nonsurgical patients. LA were measured on a digitizing computer in the 2-D apical view.

We compared patients with LA/A smoke to those without; * = $p = .0001$, AFib = atrial fibrillation:

| | # | AFib | LA + LAA THR | ↑ LA |
|---|---|---|---|---|
| + Smoke | 90 | 60 | 6 + 19 | 42 |
| | | * | * | * |
| − Smoke | 320 | 43 | 1 + 6 | 48 |

Importantly, only 12/90 cases of LA/A smoke identified by TEE were seen by TTE. By multivariate analysis only smoke correlated significantly with the presence of LA/A thrombus.

Patients with LA/A smoke are at higher risk for thrombus. Also, AFib, large atria, and smoke may constitute a dangerous triad promoting LA/A thrombus formation.

**9. The Stroke Potential of Aortic Plaque.** EA Fisher, JA Stahl, JH Budd, RL Taikowski, AS Klapper, J Weinberg, S Tuhrim, DR Horowitz, M Ostrow, ME Goldman, Division of Cardiology, Department of Medicine, Box 1030, Mount Sinai School of Medicine, 1 Gustave L. Levy Place, New York NY 10029, (212) 241-3078.

Cardiac evaluation of stroke or TIA patients includes clinical examination, ECG, and transthoracic 2-D echo (TTE). Recently, transesophageal echo (TEE) has revealed ascending aortic (AA) mobile thrombus or plaque which may be an embolic source not accessible by routine TTE.

To determine the incidence of aortic disease and its significance, we reviewed 200 consecutive nonsurgical TEE studies (patient age 60.4 ± 14 years). The entire thoracic aorta was imaged in each patient.

Fifth-two patients (26%) had aortic pathology: abnormal atherosclerotic plaque in 42 (descending in 33, ascending in 6, arch in 4), dilation in 14, dissection in 3, abscess in 1. Males had more aortic pathology (34%) than females (19%) ($p < .02$). Importantly, patients with documented recent stroke had a greater incidence of aortic pathology (31/76, 41%) than nonstroke patients (21/124, 17%) ($p = .0003$). The 6 patients with AA disease had mobile, frond-shaped masses, not detected by CT scan: 5 had central and 1 a peripheral embolism.

Thus, thorough thoracic aortic imaging with TEE may be an important component of the comprehensive evaluation of potential sources of peripheral and central emboli.

**10. Supply:Demand Ratio Predicts the Extent of Myocardial Dysfunction Observed during PTCA.** M Cohen, DH Israel, Division of Cardiology, Department of Medicine, Box 1030, Mount Sinai School of Medicine, New York, NY, 10029, (212) 241-5849.

The extent of myocardial dysfunction observed during sudden coronary occlusion (occl) may be a function of collateral blood supply versus oxygen demand (developed systolic tension). The distal coronary occlusion pressure (DCOP) measured during angioplasty has been used to estimate the adequacy of collateral flow, but fails to account for oxygen demand.

Therefore, we related the oxygen supply:demand ratio to systolic function during ischemia in 24 patients with left anterior descending or right coronary lesions and normal left ventriculograms undergoing coronary angioplasty. DCOP and LV pressure were simultaneously recorded during balloon inflation, and the area between DCOP and the LV diastolic pressure curves, the diastolic pressure time index (DPTI), served as an indicator of collateral flow. The area under the LV curve in systole, the systolic pressure time index (SPTI), served as an indicator of oxygen demand. An RAO left ventriculogram was performed during the 4th balloon inflation, and contractile dysfunction was measured as the percent of the segment at risk becoming asynergic (%ASYNR) during occlusion.

There was a strong relation between DPTI/SPTI and %ASYNR, $r = -.85$. All 8 patients with DPTI/SPTI <.3 had %ASYNR >75%. Peak systolic pressure multiplied by heart rate did not correlate with %ASYNR. The correlation between DCOP alone and %ASYNR was weak ($r = -.60$), but when 6 patients with increased demand (systolic pressure >110 or heart rate >90) were excluded, a stronger relation between DCOP and %ASYNR was observed ($r = -.80$).

Thus the supply:demand ratio is the best predictor of ischemic myocardial dysfunction after coronary occlusion. The distal coronary occlusion pressure is a simpler parameter to obtain, but is less predictive of ischemic myocardial dysfunction during angioplasty in patients with higher oxygen demand.

**11. Can the Left Ventricular Cavity/Myocardium Count Ratio in Exercise SPECT Tl-201 Scintigraphy Identify High-Risk CAD Patients?** RR Roberti, MA Baruchin, A Van Tosh, SF Horowitz, Division of Cardiology, Department of Medicine, Beth Israel Medical Center, First Ave. at 16th St., New York, NY, 10003, (212) 420-2092.

We have previously reported that a cavity:myocardium count ratio (C/M ratio) computed from exercise SPECT Tl-201 scintigrams correlates with left ventricular ejection fraction.

To test the ability of this descriptor to identify patients with coronary artery disease (CAD), we determined the C/M ratio in 115 patients who had undergone cardiac catheterization and exercise thallium scintigraphy. C/M ratio was also determined in a group of 13 patients with a low probability of CAD. The C/M ratio is calculated from the short axis midventricular slice view by dividing the counts in a 2 × 2 pixel area in the left ventricular cavity by counts in the hottest area of myocardium of the exercise thallium image.

Results: In the group of patients who underwent cardiac catheterization, 16 patients were normal and 99 had coronary disease. There was no significant difference between the average C/M ratio in the 16 normal cardiac catheterization patients and 13 patients with low probability of CAD (0.68 ± .12 vs 0.70 ± .12, $p$ = NS). The mean C/M ratio in the 29 patients with normal cardiac catheterization or low probability of cad was significantly higher than the C/M ratio in the 99 patients with CAD (0.69 ± .12 vs 0.42 ± .12, $p < .01$). The mean C/M ratio in 66 patients with CAD and normal ejection fraction was significantly lower than the C/M ratio in the normal group (0.49 ± .15 vs 0.69 ± .12, $p < .01$). Seventy patients with multivessel disease had a mean C/M ratio lower than the 29 patients with single vessel disease (.39 ± 17 vs .47 ± 17, $p < .01$). When multivessel disease patients were analyzed, 25/30 (83%) of those with EF ≤50% had a C/M ratio ≤.40, and 27/40 (68%) with EF >50% had a C/M ratio >.40.

Mean cavity/myocardium (c/m) values are presented below (CAD, coronary artery disease; EF, ejection fraction; MVD, multivessel disease):

| | | | CAD | | MVD |
|---|---|---|---|---|---|
| | All N1† | All CAD | EF > 50% | All MVD | EF ≤ 50% |
| N | 29 | 99 | 66 | 70 | 33 |
| C/M | .69 ± 12 | .42 ± 12* | .49 ± 12* | .39 ± 17* | .27 ± 14* |

† 13 low CAD probability + 16 normal angiogram patients.
* $p < .01$ compared to N1

Conclusion: C/M ratios derived from exercise SPECT thallium-201 myocardial perfusion scintigraphy is lower in coronary artery disease patients than in normal subjects, independent of resting ejection fraction. Low C/M ratio (≤.40) is a marker of more severe coronary disease and can identify high-risk patients in whom surgery may be beneficial.

**12. Clinical Predictors of SPECT Thallium-201 24-Hour Redistribution.** IL Schwartz, A Van Tosh, RR Roberti, SF Horowitz, Division of Cardiology, Department of Medicine, Beth Israel Medical Center, First Avenue at 16 Street, New York, NY 10003, (212) 420-2092.

Delayed imaging of exercise SPECT Tl-201 scintigraphy at 24 hours improves sensitivity for detecting reversible myocardial ischemia but is a time-cost burden to imaging labs.

We reviewed the clinical history, exercise stress test results, and thallium scintigraphy in 102 patients who had exercise, 4-hr and 24-hr delayed imaging to identify factors that predict 24-hour redistribution. Variables analyzed were history of angina or MI, angina on stress test, Q waves on resting ECG, and positive exercise ECG stress test. Exercise and 4-hr and 24-hr images were scored for perfusion of the anterior, septal, inferior, lateral, and apical segments using a 0-3 scale. Significant redistribution was defined as improvement of one perfusion grade in any segment from exercise to four hours, or from 4 hours to 24 hours.

Results: 41 of 102 patients showed significant 24-hour redistribution; 34/69 patients with significant 4-hour redistribution showed significant additional 24-hour redistribution compared with only 7/33 patients without 4-hour redistribution ($p < .02$). Sixteen (of 36) patients with a positive exercise ECG stress test showed significant additional 24-hour redistribution as compared to 7/52 with negative ECG stress test ($p = .003$). Ten (of 42) patients with Q waves on resting ECG had significant 24-hour redistribution as compared to 22/53 patients without Q waves (p = .10). A prior history of angina or MI, and angina during stress testing were not significantly predictive of 24-hour redistribution. Clinical variables can also be combined to create a profile of individual patients who are likely to demonstrate 24-hour redistribution. For example, 8/15 patients with the combination of significant 4-hour redistribution, a positive exercise stress test, and no Q waves on resting ECG had significant additional 24-hour redistribution. Only

3/17 patients without 4-hour redistribution, negative ECG stress test, or Q waves on rest ECG had significant 24-hour redistribution ($p < 0.04$).

Conclusions: Combinations of clinical variables can identify patients who are at high or low probability for demonstrating significant additional thallium redistribution at 24 hours. The clinical application of this analysis may improve laboratory efficiency for detecting myocardial ischemia by thallium scintigraphy.

**13. Isovolumic Relaxation Flow during Right Ventricular Pacing.** E Kirschner*, SR Hecht, M Berger, Division of Cardiology, Department of Medicine, Beth Israel Medical Center, First Avenue at 16 Street, New York, NY 10003, (212) 420-2092.

Although intraventricular flow during LV isovolumic relaxation "isovolumic relaxation flow" (IRF) has been detected by Doppler echo, this finding has been limited to pts with normal to hyperdynamic LV function and cavity obliteration.

We now describe this finding in a series of pts undergoing RV pacing. M-mode, 2D, and Doppler studies were performed in 28 pts (mean age 78 ± 13 yrs) with transvenous pacemakers. During pacing, IRF was present in 27/28 pts. Only 3/10 pts with intermittent pacing demonstrated IRF during nonpaced beats ($p < 0.001$). Flow was apically directed in 22 and bidirectional in 5. The isovolumic relaxation period was longer in pts with bidirectional flow (188 vs 126 ms) ($p < 0.001$). Peak velocity by continuous-wave Doppler ranged from 35–132 cm/s (mean 69 ± 25). IRF was localized by pulsed-wave Doppler to the level of the proximal or midseptum. Fractional shortening (FS), measured in 21 pts, ranged from 18% to 48% (mean 36% ± 7). Three pts had reduced LV function (FS < 28%) and increased LV end-diastolic dimensions (>5.6 cm). Cavity obliteration was present in only 4 pts.

Conclusions: IRF is a consistent finding during RV pacing. In these pts, its presence is independent of cavity obliteration. It may also occur with LV dysfunction and increased chamber size. We postulate that an intraventricular pressure gradient generated by an abnormal sequence of relaxation induced by RV pacing may be responsible for this finding.

**14. Angiotensin II Stimulates γ-Interferon Production: A Potential Explanation for the Beneficial Effect of Converting-Enzyme Inhibitors in Myocarditis.** Alan Lotvin, Lloyd Mayer, Henry Krum, Jill Kalman, Milton Packer, Mount Sinai School of Medicine, New York, NY 10029.

---

## Clinical Immunology

**15. Hydroxychloroquine Inhibits Release of IL-1 and IL-6 from Monocytes.** K Sperber, H Quraishi, V Stecher,* L Mayer, Division of Clinical Immunology, Box 1089, Mount Sinai Medical Center, New York, NY 10029, (212) 241-7351, and *Sterling Pharmaceuticals, New York, NY, 10017.

We examined the effect of hydroxychloroquine (HQ) on IL-1 and IL-6 production from LPS and gamma interferon stimulated monocytes. Monocytes were pretreated with serial 10-fold dilutions of 1 mM HQ and stimulated with LPS (10 μg/mL) or gamma interferon (100 U/mL) for varying periods of time (4–12 hours). Supernatants from these cells were then analyzed for IL-1 and IL-6 by both ELISA and bioassay. Pretreatment with 1 mM to $10^{-2}$ mM of HQ resulted in a 10-fold reduction in detectable IL-1 and IL-6 in culture supernatants (13 ng to 0.1 ng for IL-6, and 1 ng to undetectable levels for IL-1) by ELISA and bioassay.

These data could not be explained by alterations in viability, as no significant reduction in cell viability was observed between untreated and treated (viability 90% vs 87%) monocytes as assessed by PI staining. In order to determine whether HQ was inhibiting production of these cytokines, intracytoplasmic staining with anti-IL-1 and -IL-6 antibodies was performed on LPS (10 μg/mL) stimulated HQ treated and untreated cells.

Both intracytoplasmic IL-1 and IL-6 were detected in HQ treated and untreated monocytes, suggesting that HQ inhibits release but not synthesis of IL-1 and IL-6. In contrast to monocytes, treatment of PHA or anti-CD-3 mAb stimulated T cells with HQ failed to inhibit secretion of IL-2 or IL-4.

These different effects may reflect differences in intracellular concentrations of HQ. These data suggest that inhibition of IL-1 and IL-6 release may be an important mechanism of action of HQ.

**16. Treatment of Common Variable Immunodeficiency with PEG-IL2.** C Cunningham-Rundles, L Mayer, L Mendelsohn, E Sapira, Division of Clinical Immunology, Box 1089, Mount Sinai Medical Center, New York, NY 10029, (212) 241-7351.

Patients with common variable immunodeficiency (CVI) have decreased immunoglobulin levels and consequent infections. Although previous studies have suggested that the defect(s) resulting in this disease are intrinsic to the B cell, numerous reports of T-cell deficiencies have been described as well, including reduced IL-2 production. Addition of T-cell-derived cytokines to B cells in vitro has resulted in increased Ig secretion. Thus the concept of an irrevocable B-cell defect in all patients with CVI may not be valid. The Ig deficiency of CVI may reflect improper communication between defective T cells and autologous B cells, resulting in aberrant B-cell maturation in vivo.

To directly assess such a possibility, we have treated 5 CVI patients with a new biologic, human recombinant IL-2 conjugated to polyethylene glycol. Doses were 250,000 IU/m$^2$ weekly for weeks 1–4, 500,000 IU/m$^2$ for weeks 5–8, and $10^6$ IU/m$^2$ for weeks 9–12. Although lymphocyte phenotypic characteristics did not change appreciably, proliferative responses to T-cell mitogens were reduced in all cases. In contrast, secretion of Ig in vitro by CVI B cells in response to both T-cell-dependent (autologous or allogeneic T cells and pokeweed mitogen) and T-cell-independent (SAC, SAC and BCDF) stimulation was increased 100-10,000–fold.

Although these data demonstrate an improvement in B-cell responsiveness to differentiation signals, there was a concomitant improvement in T helper activity as well. Increases of 100-1000–fold in Ig secretion by normal allogeneic B cells were detected when co-cultured with in vivo PEG-IL-2 treated CVI T cells, a level comparable to normal T cells. Despite the improvement in in vitro function, no increase in serum Ig was noted throughout the study period. Modest improvement in clinical status was noted in 4/5 patients.

Whether the effects of PEG-IL2 are mediated through T cells or directly on B cells is unknown, but these data reinforce the concept that CVI B cells are competent yet may lack certain T-cell-derived signals, critical for response to maturation signals, in vivo.

**17. Purification and N Terminal Sequence of a B Cell Activation Antigen Associated with Interleukin-6 Mediated B-Cell Differentiation.** R Huang, A Pizzimenti, R Furman, M Cidon, L Mayer, Division of Clinical Immunology, Box 1089, Mount Sinai Medical Center, New York, NY 10029, (212) 241-7351.

We have previously described a monoclonal antibody, 54, generated by immunizing Balb/c mice with the B lymphoblastoid line CESS, which inhibits IL-6 induced differentiation of CESS. mAb 54 stains activated B cells, resting monocytes, Hep G2 cells, and the T-cell line Jurkat while it fails to stain resting B cells, T cells, or activated monocytes. This pattern of staining correlates with the known cell reactivity to IL-6. mAb 54 also inhibits IL-6 and pokeweed mitogen induced Ig secretion by normal peripheral blood B cells.

Since PWM stimulates IL-6 production by T cells, we wanted to determine whether 54 recognizes the IL-6 receptor, previously isolated and cloned by Kishimoto and his colleagues. mAb 54 precipitates an 80kd band from surface-iodinated CESS cells which is comparable to the known molecular weight of IL-6R. However, we were unable to pre-clear the 80kd band with the anti-IL-6R mAb, MT18. Furthermore, using phycoerythrin labelled IL-6, only mAb MT18 but not mAb 54 was capable of inhibiting binding of IL-6 to its receptor. We next purified the surface protein recognized by mAb 54 by immunoprecipitation and trasnfer onto PVDM membranes. The N terminal sequence was determined by automated sequencing. The first 34 amino acids were not related to the known sequence of the IL-6R and appear to be novel by comparison with other known human, murine, bacterial, and viral proteins.

Taken together, these data suggest that mAb 54 recognizes a cell-associated molecule which correlates with IL-6 but not other B-cell differentiation signals. Whether this protein represents a novel IL-6R associated molecule remains to be determined.

**18. Adhesion Molecules Are Expressed on Intestinal Epithelial Cells but Do Not Participate in T-Cell Interactions.** F Girardi, A Panja, P Salomon, L Mayer, Divisions of Clinical Immunology and Gastroenterology, Box 1089, Mount Sinai Medical Center, New York, NY 10029, (212) 241-7351.

Leukocyte function associated-antigens (LFA) have been clearly defined as playing a key in accessory-cell : T-cell, T-cell : B-cell, and cytolytic-T-Cell:target-cell interactions. These adhesion molecules appear to enhance low affinity interactions, allowing for appropriate cellular activation and effector functions. We have recently described a novel accessory cell in the mucosa associated lymphoid tissue, the intestinal epithelial cell, which is capable of stimulating antigen-primed and alloreactive T cells. Since epithelial cells from other sites appear to possess ligands for adhesion molecules, we wanted to determine whether intestinal epithelial cells express LFA molecules and/or their ligands. In addition we examined whether this family of integrins is critical to epithelial accessory cell function.

Antibodies to LFA-1a, -2, and -3 were used to stain frozen sections of normal intestine. mAbs against LFA-1a and -2 stained T cells in the lamina propria but failed to stain intestinal epithelial cells. Only the anti-LFA-3 mAb stained intestinal epithelium. These antibodies were then used in mixed lymphocyte cultures (MLR), where either irradiated normal epithelium or non-T cells served as the stimulator cell population.

Whereas all three mAbs inhibited conventional MLRs, none of the three mAbs inhibited epithelial cell driven MLR cultures. When epithelial-cell : T-cell MLRs were performed using lamina propria lymphocytes (LPL) as the responder population, the findings were similar.

These data suggest that despite the presence of adhesion molecules in the tissue, this family of integrins does not regulate T-cell:accessory-cell interactions in the gut. Therefore alternate pathways of T cell activation are in effect when epithelial cells serve as accessory cells. These differences may relate to their unusual function in the mucosal immune system.

**19. Induction of Human B-Cell Differentiation via a Gi-Linked Cytokine Receptor.** R Huang, M Cidon, R London, L Mayer, Divisions of Clinical Immunology and Renal, Box 1089, Mount Sinai Medical Center, New York, NY 10029, (212) 241-7351.

We have previously reported a novel human B-cell differentiation factor, 446-BCDF, which is distinct from other known cytokines (for example, IL-6) and binds to a unique receptor on activated B cells. The signal pathway appears to involve cAMP, since agents which induce a rise in cAMP result in inhibition of the response to 446-BCDF.

In order to determine the role of cAMP in this system, we activated isolated B cells with *Staphylcoccus aureus* Cowan I strain (SAC), in the presence or absence of 446-BCDF, and measured changes in intracellular cAMP at varying time points.

cAMP levels in resting or SAC-activated B cells did not change appreciably over a 24 h period (0.34pM/$10^6$ cells ± 0.09). In contrast, after exposure to 446-BCDF, intracellular cAMP concentration started to decrease at 5 min and was significantly lower at 30 min (0.38 vs 0.18pM/$10^6$ cells, $p < 0.025$) and at 4 h (0.45 vs 0.10pM/$10^6$ cells, $p < 0.025$). After 24 h, cAMP concentrations started to increase to baseline.

In order to determine whether the reduction of cAMP was related to effects on production or breakdown, we tested the effects of angiotensin II on B-cell differentiation. Angiotensin II binds to a Gi-linked receptor present on mononuclear cells which would inhibit cAMP production. Addition of angiotensin II at concentrations of $10^{-8}$–$10^{-10}$M to SAC-activated B cells resulted in the induction of Ig secretion.

These data suggest that 446-BCDF binds to a Gi-linked receptor on B cells, inhibiting cAMP and promoting Ig secretion.

**20. CD1d May Be Critical in Epithelial-Cell:T-Cell Interactions.** A Panja, R Blumberg,* S Balk,* P Bleicher,* L Mayer, Divisions of Clinical Immunology, Box 1089, Mount Sinai Medical Center, New York, NY, (212) 241-7351, and Gastroenterology, The Brigham and Woman's Hospital, Boston, MA.

CD1 is a class-l-like molecule which is expressed on thymocytes and dendritic cells and is thought to be involved in the education of T cells in the thymus. Recently, CD1 has been demonstrated on murine intestinal enterocytes, adding further speculation as to the role of the intestinal epithelium as an extrathymic region for T cell development. We have previously described the finding that normal intestinal epithelial cells stimulate CD8+ antigen-nonspecific suppressor T cells potentially through a novel pathway of T-cell activation. Since CD1 is class-l-like, this molecule may be acting as a ligand for CD8.

Using recently developed rat monoclonal antibodies to mouse and human CD1, 3C11, and 1H1, we studied the ability of anti-CD1 mAbs to inhibit epithelial-cell:T-cell mixed lymphocyte cultures. Complete inhibition of epithelial cell driven MLR responses by mAb 3C11 was seen in a dose-dependent fashion (0.1 μg/mL, 50% inhibition; 1–10 μg/mL, 100% inhibition). Lesser inhibition was detectable when mAb 1H1 was tested. In contrast, no inhibition of conventional MLRs (stimulation with irradiated non-T cells) was seen with either mAb at any concentration, suggesting that the inhibition seen with epithelial cells was not related to nonspecific inhibition of T-cell activation.

Since CD1 is expressed on the basolateral aspect of intestinal epithelial cells, we next tested the ability of these mAbs to inhibit MLR cultures where lamina propria lymphocytes served as the responder population. Similar to the results seen with peripheral blood T cells, LPL proliferation was inhibited by 3C11 in epithelial cell MLR, but no inhibition was seen when non-T cells were used as the stimulator cells. Since epithelial cells from patients with IBD fail to stimulate suppressor T cells in MLR, we evaluated the ability of anti-CD1 mAbs to inhibit in this system. Similar to the effect on normal epithelial-cell stimulated MLRs, 3C11 completely inhibited IBD epithelial cell MLRs, suggesting that the effect of 3C11 was not CD8+ T-cell specific. Although these results may reflect steric hindrance on the part of the anti-CD1 mAbs, control antibodies which bound to normal epithelial cells failed to inhibit T-cell proliferation.

Taken together, these data suggest that CD1 may serve as one possible restriction element in epithelial cell:T-cell interactions implicating an alternate pathway of T-cell activation in the intestine.

**21. Cytokine Production by Normal and Inflamed Intestinal Epithelial Cells: Accessory Cytokines for T-Cell Activation.** A Panja, E Siden, L Mayer, Division of Clinical Immunology, Box 1089, Mount Sinai Medical Center, New York, NY 10029, (212) 241-7351.

Intestinal epithelial cells are the first line of defense in mucosa-associated lymphoid tissue. We have previously demonstrated that normal epithelial cells are functional antigen-presenting cells (APC), capable of stimulating alloreactive and antigen-primed T cells.

Since accessory cytokines are critical for T-cell activation, we tested for the presence of mRNA for IL-6 and IL-1 in freshly isolated enterocytes by Northern and slot blot analysis. mRNA was isolated by either the cesium chloride density gradient or RNAzol method. mRNA for IL-6 and Il-1 were probed using a 600 kb nick translated riboprobe in a pGEM vector.

In each case mRNA for IL-6 was detected at the appropriate size (1.3 kb). However, no IL-1 mRNA was evident in these same blots after stripping and rehybridization. To determine whether IL-1 mRNA is induced during inflammation, mRNA from epithelial cells from inflamed areas of bowel in patients with Crohn's disease and ulcerative colitis were obtained. IL-6 mRNA was comparable to normal epithelium in both disease states; however, IL-1 mRNA was still undetectable. The absence of IL-1 is further evidence for the difference between enterocytes and macrophages as APCs. In contrast, TNF mRNA was detectable in IBD epithelial cells whether the bowel was actively inflamed or lacked evidence of active inflammation.

These data suggest that epithelial cells produce at least one important accessory cytokine which may be increased in response to inflammation. The secretion of IL-6 and TNF may accentuate the T- and B-cell activation in inflammatory bowel disease.

## Endocrinology

**22. Stimulation of Glut1 Glucose Transporter Gene Transcription by Thyroid Hormone in ARL15 Cells.** SP Weinstein, J Watts, RS Haber, Division of Endocrinology and Metabolism, Department of Medicine, Box 1055, Mount Sinai School of Medicine, New York, NY 10029, (212) 241-3792.

The enhancement of metabolic rate by thyroid hormone is accompanied by increased glucose uptake in target tissues as well as in thyroid hormone-responsive cultured cells. In the rat liver-derived ARL 15 cell line, we have previously shown that a greater than twofold stimulation of glucose transport by $T_3$ is attributable to a corresponding increase in the abundance of cellular glucose transporters and of mRNA coding for the Glut1 (erythroid) glucose transporter isoform. We have now further characterized the mechanism by which $T_3$ increases glucose transport in these cells.

Treatment of confluent monolayers of ARL 15 cells with 50 nM $T_3$ clearly increased both the rate of $[^3H]$2-deoxyglucose uptake and Glut1 mRNA abundance by as early as 4 hours. The maximal effect at 48 hours was 2-3–fold for both parameters. The half-life of the Glut1 mRNA, calculated from its disappearance rate in the presence of actinomycin D, was not detectably changed by $T_3$ treatment for 48 hours (controls, 3.3 ± 0.7 hours; $T_3$-treated, 3.7 ± 1.0 hours; $n = 4$ experiments). In these experiments $T_3$ increased the abundance of the Glut1 mRNA by 92 ± 17%. The same $T_3$ treatment protocol increased the rate of transcription of the Glut1 gene, determined by nuclear run-on analysis, by 50 ± 16%, normalized to β-actin, whose transcription was not increased ($n = 3$ experiments). To determine whether the induction of the Glut1 mRNA might be a direct effect of $T_3$ on Glut1 gene transcription, we assessed the sensitivity of this $T_3$ effect to protein synthesis inhibitors. Exposure of cells for 6 hours to $T_3$ alone increased Glut1 mRNA by 43 ± 10%. In the presence of cycloheximide ($4 \times 10^{-5}$ M) or anisomycin ($10^{-4}$ M) the Glut1 mRNA in control cells increased by 5- and 7-fold, respectively, at 6 hours; there was no further induction of this mRNA by $T_3$ in the presence of these inhibitors.

We conclude that $T_3$ regulates cellular glucose transport, at least in part, via increased glucose transporter gene transcription. Whether this effect is direct, or instead represents an adaptive response to enhanced cellular glucose utilization, is not clear at present.

**23. A Splicing Variant of Rat Thyroglobulin mRNA (rTg-2) Is Detected only in the Presence of TSH.** PN Graves, TF Davies, Division of Endocrinology, Department of Medicine, Mount Sinai School of Medicine, New York, NY 10029.

We recently described a 1kb thyroglobulin transcript (rTg-2 mRNA) in cultured rat thyrocyte lines and rat thyroid tissue which shared the first 5 exons of 9kb rTg-1 mRNA and contained additional 3′ coding and noncoding sequences. Using the cultured rat thyroid cell model, the levels of both rTg mRNAs were similar in the presence of TSH. However, <5% of rTg-2 mRNA was detectable by blot hybridization 24 hours after removal of TSH while >60% of rTg-1 mRNA remained. These data indicated differential mechanisms of TSH control and raised the possibility of two separate Tg genes.

A genomic clone, therefore, was isolated which contained the entire rTg-2 mRNA sequence. The promoter sequence was identical to the published rTg-1 promoter, consistent with alternative splicing of a single primary Tg transcript.

These data imply that TSH increases rTg-2 mRNA by enhancing alternative splicing and concomitant cleavage/polyadenylation. The rapid disappearance of rTg-2 mRNA following TSH removal indicated an unstable mRNA structure. A single AUUUA destabilizing element adjacent to the poly(A) tail present in several rTg-2 cDNA clones was altered by absence of 2 or 3 bases in several other clones. This element was absent from rTg-1 mRNA. PCR-amplified genomic DNA contained only TAAAT in the coding strand, indicating that the alteration occurred at the RNA level and that turnover involves both endo- and exonuclease activities directed at this target sequence.

These data demonstrate that TSH, as the primary trophic hormone of the thyrocyte, regulates expression of the Tg gene at multiple sites and offers a unique model for exploring mRNA stability.

**24. Intrahypothalamic GT1 Cells Differentiate and Stimulate Reproductive Development in Hypogonadal Mice.** MJ Gibson, AJ Silverman, JL Roberts, KW Dong, GM Miller, Division of Endocrinology, Department of Medicine, and Fishberg Center for Neurobiology, Box 1055, Mount Sinai School of Medicine, New York, NY 10029, (212) 241-7544, and Columbia University, New York, NY 10032.

Bilateral injections of GT1-7 cells were made into the anterior medial hypothalamus of adult mutant hypogonadal (hpg) mice, which fail to produce GnRH. GT1 cell lines are immortalized neurons, developed by genetically targeted tumorigenesis, that produce and secrete GnRH in culture.

Neuronal-appearing cells with a plexus of fibers innervated the host brain at 1 wk. Cells were immunoreactive to GnRH antibodies LR1, 1076 (which detect precursor and product), and EL14 (which reacts with product only), suggesting complete processing of the GnRH prohormone in grafted cells. Numerous differentiated cells were seen in the hpg host brain at 5 wk.

Significant gonadal development in response to GT1-7 cells was evident in all 3 male hpgs studied 10 wk after grafting, with testes weights of 19.2, 20.4, and 33.8 mg (vs. untreated 7.6 mg). Two hpg female mice responded to GT1-7 implants with vaginal opening within 1 wk, reflecting ovarian stimulation; at 10 wk one hpg had significant uterine development (36.0 mg).

We conclude that GT1 implants in hpg mice provide a powerful model to evaluate the utility of cell lines for neuronal transplants.

Supported by: NS20335 (MJG) & HD19077 (MJG).

**25. Effect of Androgen, Estrogen, and Growth Factors on the Proliferation of Cultured Fibroblasts Derived from Human Fetal and Adult Prostates.** AC Levine, M Ren, G Huber, A Kirschenbaum, Division of Endocrinology and Department of Medicine, Box 1055, Mount Sinai School of Medicine, New York, NY 10029, (212) 241-6095.

Prostatic stroma is felt to play an important role in fetal prostatic epithelial-cell differentiation as well as in the development of benign prostatic hypertrophy (BPH). These stromal-driven processes are both under the regulation of androgens and possibly estrogens. We have isolated prostatic fibroblasts from both adult (suprapubic prostatectomy specimens) and fetal (19–20 week abortion) tissue via a collagenase method and tested the effects of growth factors and sex steroids on cellular proliferation.

Cells from passages 3–7 were grown in Ham's F12 media alone (serum-free) or Ham's F12 with 1% charcoal-stripped (cs) serum. Estradiol ($E_2$), dihydrotestosterone (DHT), basic fibroblast growth factor (bFGF), and epidermal growth factor (EGF) were added daily and proliferative effects assessed on day 6 via cell counts (hemocytometer) and $^3$H-thymidine incorporation. Both EGF and FGF stimulated adult and fetal prostatic cell growth in a dose-dependent fashion. Estradiol had no effect on cellular proliferation. DHT ($10^{-8}$ M and $10^{-7}$ M) caused a small but significant ($p < .05$) enhancement of fetal fibroblast proliferation in serum-free media only. DHT did not enhance fibroblast cellular proliferation if the media was changed daily. Immunohistochemical staining for estrogen receptors was negative. However, nuclear androgen receptors (AR) were weakly demonstrable.

Fetal fibroblasts were deprived of serum for 3 days and then probed for the presence of the mRNA for the proto-oncogene c-fos after 20 min stimulation with either serum, bFGF, or DHT. Both serum and bFGF induced the mRNA for c-fos, whereas DHT had no effect.

We conclude that estradiol has no effect on prostatic fibroblast proliferation. DHT enhances prostatic fibroblast proliferation only in serum-free media. This effect appears to be indirect, because DHT does not induce the mRNA for c-fos. Both EGF and bFGF enhance human prostatic fibroblast proliferation. Androgens may enhance the production and secretion of growth factors from prostatic stromal cells, which then stimulate prostate growth in an autocrine and paracrine fashion.

**26. Intrathyroidal T-Cell Receptor Vα rather than Vβ Gene Restriction in Human Autoimmune Thyroid Disease (AITD).** TF Davies,* A Martin, P Graves, N Lahat, WL Cohen, A Ben-Nun, Depts. of Medicine, Mount Sinai School of Medicine, New York, NY, Lady Davis Carmel Hospital, Haifa, and Department of Cell Biology, Weizmann Institute of Science, Rehovot, Israel.

To define the nature of preferential intrathyroidal lymphocyte (ITL) accumulation we have investigated human T-cell receptor (hTcR) gene family utilization within the glands of patients with AITD.

We transcribed cDNA from cellular RNAs, obtained from intact thyroid tissue and from T-cell cultures of ITL, and performed 35-cycle polymerase chain reactions (PCRs) with specific oligonucleotides to amplify 18 Vα gene and 19 Vβ gene families. Peripheral blood (PBMC) and benign thyroid adenomas provided control RNAs. All PCR products were hybridized with Cα or Cβ oligoprobes to confirm specificity.

We found that most Vα and Vβ gene families were detected in cDNAs from PBMC ($n = 6$), thyroid adenoma tissue ($n = 8$) and normal human thyroid. There was a similar pattern of Vβ gene family usage by ITL cultures and/or thyroid tissue from 10 patients with Graves' disease or Hashimoto's thyroiditis, with an average of 14.4 out of the 19 V$\beta$ gene families easily detected in each sample. In contrast, utilization of V$\alpha$ genes was markedly reduced within both ITL T-cell populations as well as within intact tissue from similar patients (mean 4.0 out of 18 families easily detected, $p < 0.01$). There was no consistent pattern in the V$\alpha$ gene restriction, although all 10 samples analyzed utilized V$\alpha$1, V$\alpha$11, or V$\alpha$12. Such a pattern in V$\alpha$ gene family usage was not mimicked in 7-day autologous mixed lymphocyte reactions ($n = 6$).

These data imply specific hTcR V$\alpha$ gene utilization within the thyroidal environment of human AITD. Such V-gene activation may be secondary to multiepitopic thyroidal autoantigens or to the presence of a thyroidal antigen capable of activating a variety of T cells with restricted V$\alpha$ genes, a thyroidal "superantigen."

**27. Early Screening for Gestational Diabetes Is Predictive of Insulin Use.** M Alvarez-Marfany, A Stagnaro-Green, M Alvarez, SH Roman, Division of Endocrinology and Metabolism, Departments of Medicine and Obstetrics and Gynecology, Box 1055, Mount Sinai School of Medicine, New York, NY 10029, (212) 241-8797.

The identification and adequate therapy of women with gestational diabetes (GDM) is of recognized importance to fetal outcome. At The Mount Sinai Hospital prenatal clinic, women are screened for diabetes at their initial visit regardless of gestational age. Patients who initially have negative tests are rescreened at 28 weeks of gestation or before if indicated. The aim of this retrospective study was to determine the effect of early screening for GDM on diagnosis and treatment.

Thirty women (mean age 30.5 years, range 17–44 years) who developed GDM during their pregnancy and were initially screened prior to 24 weeks (mean initial OGTT at 13 weeks, range 6–23) were included. Eighteen women (60%) were found to have GDM before 28 weeks (group 1), and 12 women (40%) were diagnosed after 28 weeks (group 2). The mean week of diagnosis was 16.3 weeks for group 1 and 32.6 weeks for group 2 ($p = 0.0001$). Insulin was required to achieve adequate control of blood glucose levels in 12/18 (66%) of women from group 1, in contrast with 3/12 (23%) of women from group 2, $\chi^2 = 5$, $p = 0.025$.

| | Insulin | No Insulin |
|---|---|---|
| Before 28 ws | 12 | 6 |
| After 28 wks | 3 | 9 |

$\chi^2 = 5.7 \quad p = 0.017$

Mean gestational week of initiation of insulin was 23.4 in the group diagnosed early as compared with 32.6 in the women diagnosed at 28 weeks or later, $p < 0.05$. The observed mean difference of prepregnancy body mass indices between the two groups was not significantly different.

We conclude that (1) GDM can be diagnosed in the majority of pregnancies prior to 28 weeks, (2) early diagnosis of GDM identifies women who will likely require insulin during pregnancy, (3) early- and late-onset GDM may represent two distinct patient populations.

**28. Gaba-B Mediation of NMDA-Stimulated LH Release in Female Mice.** GM Miller, MJ Gibson, Division of Endocrinology, Department of Medicine, Box 1055, Mount Sinai School of Medicine, New York, NY 10029, (212) 241-7686.

Our recent findings suggest central inhibitory mediation of GnRH secretion to NMDA challenges, resulting in differential release of LH in male and female hypogonadal mice with preoptic area brain grafts (HPG/POA). As GABA blocks NMDA-induced LH release in pubertal rats, we evaluated the role of the GABA-B receptor in modulating NMDA effects in normal female mice.

Five days after ovariectomy, mice were fitted with jugular cannulae and tested the following day. Two NMDA challenges (20 mg/kg, IV) were given 2 hr apart. Ten min before the 2nd NMDA challenge, mice received the GABA-B receptor agonist baclofen (Bac, 5 mg/kg, IV) or saline (Sal). Blood samples (0.1 mL) were taken at 10-min intervals before and after each challenge. Resuspended red blood cells were returned following the next sample withdrawn, and plasma duplicates (25 $\mu$L) were assayed for LH.

Bac-pretreatment attenuated the LH response to the 2nd NMDA challenge: 1 : 2.21 ± 0.36 ng/mL, $p < 0.01$; 2 : 1.28 ± 0.22 ng/mL, ns. In Sal-treated mice, LH increased significantly to both NMDA challenges: 1 : 3.04 ± 0.38 ng/mL, 2 : 3.06 ± 0.38 ng/mL, both $p < 0.01$. HPG/POA female mice are being tested.

These studies implicate GABA-B receptors in GABAergic control over NMDA-induced LH release in female mice.

Support: NIH Grants T32DK07645 (GM) & HD19077 (MJG).

**29. GnRH Cells in the Hypogonadal Mouse Form Functional Connections Despite Their Biosynthetic Deficiency.** I Livne, MJ Gibson, AJ Silverman, Dept. of Medicine, Mount Sinai School of Medicine, New York, NY 10029, & Dept. of Anatomy & Cell Biology, Columbia University, New York, NY 10032.

The hypogonadal (hpg) mouse has an undeveloped reproductive tract due to a truncated GnRH gene, resulting in failure to synthesize GnRH. Cells in the OVLT region of the hpg brain have low levels of GnRH mRNA but are devoid of the neuropeptide. The normal distribution of GnRH transcribing cells suggests that in spite of their failure to synthesize the peptide, GnRH cells migrate properly to their residence areas in the CNS. To elucidate the extent of differentiation of the mutant GnRH cells we have performed an in situ hybridization study in mice which were given injections of the fluorescent tracer Fluoro-Gold (FG). Only cells terminating on fenestrated capillaries, such as those of the hypophyseal portal system in the median eminence (ME), can capture and concentrate systemically delivered FG. We have therefore hypothesized that the finding of FG in cells containing GnRH mRNA will indicate that GnRH cells in the hpg brain are capable of innervating their proper target.

The experimental procedure for detection of GnRH mRNA and FG in the same cell was performed as follows: Normal and hpg adult males were given 3 IP injections of 40 mg/kg FG and perfused. Sections through the septal-POA-AHA extent were processed for standard in situ hybridization with [$^{35}$S] labeled oligonucleotide complementary to the GnRH coding region. The slides were analyzed to identify GnRH transcribing cells which accumulate FG.

In brain sections from 2 hpg males, 54.4% and 64% of the GnRH transcribing cells were double labeled with FG, while in 1 normal brain 61.2% were double labeled.

We have demonstrated that GnRH transcribing cells in the hpg brain concentrate FG. This finding suggests that mutant GnRH cells are capable of elaborating functional connections with their major target, the ME, in the absence of their principal secretory product.

NS 20335.

**30. Regulatory Factors that Influence Class II Expression in Activated T Cells.** Nancy D. Urban, Karen S. Zier, Division of Endocrinology, Mount Sinai School of Medicine, Box 1055, Mount Sinai School of Medicine, New York, NY 10029.

The MHC class II genes exhibit a tissue-specific distribution, the expression of which may be regulated at several levels. Our interest has been in the control of class II expression in activated T cells. Several DNA-binding factors have been cloned which bind to discrete regions in the 5′ promoter region. The complexity of control is suggested by the observation that the presence of these factors has been detected in class II positive and negative cells. hxbp-1, which binds to the X box, has been proposed to exert a positive regulatory role on class II regulation, since the level of its expression is reduced in class II deficient malignant B cells. YB-1, which binds to the Y box, has been proposed to play a negative regulatory role in class II regulation, since there is an inverse relationship between YB-1 levels and class II expression in certain malignant cell lines.

In order to examine whether these two factors were involved in class II regulation in normal T cells, T cells were stimulated with PHA and expanded in medium containing IL-2. RNA was isolated from aliquots of these cells on a daily basis. Following separation on a 1% agarose gel and transfer to a nylon filter, we probed with $^{32}$P-cDNAs for DR beta, hxbp-1, and YB-1.

Our results demonstrated that DR beta and YB-1 mRNA were both strongly expressed on days 1 and 2, at which point expression began to decrease in parallel. In contrast, hxbp-1 was strongly expressed from days 1 to 6.

These results are consistent with a role for hxbp-1 in the positive regulation of class II in T cells, but not negative regulation by YB-1. Moreover, they suggest that the role of nuclear binding proteins in the regulation of class II in activated T cells is complex and supports the hypothesis that they may function through as-yet-unidentified intermediate products.

**31. Positive Regulation of Fetal Human TSH Receptor mRNA by Recombinant Human TSH (rec-hTSH).** GK Huber, PN Graves,* ES Conception, TF Davies, Division of Endocrinology, Mount Sinai School of Medicine, New York, NY 10029.

Low concentrations of TSH have been shown to up-regulate the TSH receptor (TSHR), both in vitro and in vivo, while higher concentrations desensitize and down-regulate the TSHR. However, recent reports have shown only TSH-induced down-regulation of TSHR-specific mRNA transcripts with no evidence for mRNA induction. We have investigated the regulatory effects of rec-hTSH (Genzyme Inc., Framingham, MA) on TSHR-specific mRNA in human fetal thyroid monolayer cells to further explore the dose-related effects of TSH in an autologous system.

Human fetal thyroid cells were deprived of TSH for 5 days and subjected to kinetic studies with increasing concentrations of rec-hTSH in serum-free medium. Total cellular RNA was extracted and TSHR-specific mRNA analyzed using a TSHR-specific DNA probe. Low levels of a 4.3 kb human TSHR-specific mRNA were detectable after 5 days' TSH deprivation and were markedly increased by the addition of rec-hTSH ($10^3$ mU/L). This increase was detectable after only 6 hr and reached a maximum 36.1 ± 1.7–fold increase after 48 hr as judged by Northern blot analyses. Higher concentrations of rec-hTSH ($>10^4$ mU/L) had no further stimulatory effects on TSHR-specific mRNA transcript levels. The regulation of TSHR-specific mRNA was then examined after stimulation with $10^3$ mU/L rec-hTSH for 48 hr and followed by the addition of high concentrations of rec-hTSH (up to $10^5$ mU/L) for 24 hr.

These studies resulted in a loss of the rec-hTSH–induced positive regulation of TSHR mRNA (arbitrary OD units: 0.95 ± 0.2 compared to 10.5 ± 2.5).

We conclude that physiological concentrations of rec-hTSH induce the accumulation of TSHR-specific mRNA transcripts in human fetal thyroid cell monolayers. These findings are in contrast to earlier studies with immortalized rat thyroid cells and TSHR gene transfected cell models. The TSH receptor gene may, therefore, exhibit unique regulatory elements when functioning in normal human thyroid cells.

**32. Highly Purified T Cells Cannot Be Induced to Proliferate via Interaction with Immobilized Anti-CD3 Monoclonal Antibody.** Shara BA Cohen, Karen S Zier, Division of Endocrinology, Department of Medicine, Box 1055, Mount Sinai School of Medicine, New York, NY 10029.

Although it is known that monoclonal antibodies (Mab) directed to the CD3 surface molecule will mediate polyclonal T-cell activation, there has been some controversy concerning the requirement of accessory cells, or exogenous factors, when anti-CD3 Mab are immobilized by binding to a solid phase. By comparing various anti-CD3 Mab and protocols of Mab immobilization, we have confirmed that highly purified human T cells are unreactive to anti-CD3 antibody alone.

We initially compared the ability of the Mab 64.1, Leu4, and OKT3 to induce proliferation and showed that these Mab were able to stimulate peripheral blood lymphocytes (PBL) to proliferate in a specific hierarchy: 64.1 > Leu4 > OKT3; however, none of these antibodies could stimulate purified T cells to proliferate. We then attempted to induce purified T cells to proliferate by altering the immobilization protocol.

The greatest proliferative response of PBL was achieved when 64.1 Mab was immobilized to plates in RPMI without HEPES (compared with RPMI with HEPES or 50 mM Tris pH 9.5). Although PBL proliferated in response to immobilized anti-CD3 antibody alone, neither 1 µg/well nor 4 µg/well of the 64.1 Mab in any of the three media could stimulate the highly purified T-cell population to proliferate. Indeed, our T cells only responded to the immbolized anti-CD3 Mab when other modulators were used, such as in conjunction with an anti-CD28 Mab or with the lymphokines IL-2, IL-4, IL-6, or IL-7. None of these modulators had a proliferative effect on the purified T cells when used alone.

Our results confirm and extend the findings of others, in that immobilized anti-CD3 Mab cannot induce highly purified T cells to proliferate. However, flow cytometric analysis indicated that the purified T cells were activated by interaction with the anti-CD3 Mab (Tac expression was detected within 48 hr of anti-CD3 stimulation). The synergism between the signals supplied by immobilized anti-CD3 antibody and other modulators implies that there are at least two intracellular pathways involved in activating T cells to proliferate.

**33. HLA-DNA (DZα) and HLA-DOB (DOβ) Gene Expression in Human T Cells.** SBA Cohen, H Caplan, S Salvadori, KS Zier, Division of Endocrinology, Department of Medicine, Box 1055, Mount Sinai School of Medicine, New York, NY 10029.

The class II region of the human leukocyte antigen complex contains 13–15 genes (depending on the haplotype) and codes for DR, DQ, and DP molecules. It is well established that human T cells express these HLA class II molecules following activation. In addition to these functional class II genes, HLA-DNA (DZα) and HLA-DOB (DOβ) genes have been described. While it has been reported that the mRNA of these genes is transcribed, no corresponding protein product has been identified.

In order to determine whether HLA-DNA or DOB genes are expressed in human T cells, we have analyzed the HLA-DNA and HLA-DOB transcription, and mRNA steady-state levels following stimulation.

Using nuclear run on assays we obtained evidence that the DNA gene, but not the DOB gene, is transcribed in resting and activated T cells. Furthermore, Northern blot analysis demonstrated the presence of steady-state HLA-DNA mRNA (but, as expected, no DOB) in activated human T cells, and these steady-state levels increased with activation.

The presence of DNA and absence of DOB mRNA demonstrates that these class II products are not coordinately expressed in T cells. Thus, it seems unlikely that the DNα and DOβ chains constitute a heterodimer, as previously suggested. Furthermore, our results imply that expression of HLA-DNA in human T cells is regulated, at least in part, post-transcriptionally.

**34. Epidemiology of Diabetic Ketoacidosis in a Large Metropolitan Hospital.** RS Watts, M Alvarez-Marfany, A Stagnaro-Green, SH Roman, Division of Endocrinology and Metabolism, Department of Medicine, Box 1055, Mount Sinai School of Medicine, New York, NY 10029, (212) 241-8797.

We performed a retrospective chart review of admissions for diabetic ketoacidosis (DKA) at The Mount Sinai Hospital in New York City from 1/89 to 12/89. Fifty-two admissions met the following criteria for DKA: (1) glucose >200 mg/dL, (2) bicarbonate <21 mEq/L or arterial pH >7.36, and (3) serum acetone >1+.

The incidence of DKA was 0.15% (52/35, 259). The racial breakdown of our patients included 53.8% (n = 28) blacks, 34.6% (n = 18) hispanics, and 11.5% (n = 6) whites. There were 7 (13.5%) instances of new-onset diabetes (mean age 43.8 years, range 16 to 80 years), 40 (76.9%) patients with a previous history of type 1 diabetes (mean age 33.2 years, mean duration of disease 13.3 years), and 5 (9.6%) patients with a previous history of type 2 diabetes (mean age 51.4 years, mean duration of disease 1.6 years). There were no significant differences between groups for blood glucose, acetone, pH, or anion gap. During treatment of DKA, hypoglycemia occurred in 21.2% and hypokalemia in 28.8% of patients. Precipitating events were infection in 40.4% (n = 21), noncompliance in 26.2% (n = 14), infection with noncompliance in 11.5% (n = 6), miscellaneous (e.g., gastrointestinal bleeding, uremia, change in insulin regimen) in 5.8% (n = 3), and unknown in 15.4% (n = 8). There was one death unrelated to DKA, an overall mortality rate of 1.9%.

We conclude that (1) the death rate of DKA at The Mount Sinai Hospital of 1.9% was below the national average of 7%, (2) DKA as a presentation of type 1 diabetes in adults is not uncommon, (3) individuals with a previous history of type 2 diabetes can present in DKA, and (4) infection and/or noncompliance are the most frequent precipitating causes of DKA (78% of all the cases).

**35. T Cells from Adult-Onset Type 1 Diabetics Do Not Evidence the IL-2 Defect Seen in Classical Type 1 Diabetics.** RS Watts, SH Roman, A Stagnaro-Green, KS Zier, Division of Endocrinology and Metabolism, Department of Medicine, Box 1055, Mount Sinai School of Medicine, New York, NY 10029, (212) 241-8797.

Insulin-dependent (type 1) diabetes mellitus (IDDM) is believed to result from an autoimmune attack on the insulin-producing pancreatic beta cells. Studies in humans and in animal models have strongly implicated T cells in pancreatic beta-cell destruction. Previously, we reported that T cells from children with classical IDDM produced decreased levels of IL-2 following activation with phytohemagglutinin (PHA), independent of their metabolic control or their duration of disease. However, it is becoming more apparent that IDDM is a heterogeneous disease, and the role that autoimmune phenomena play in some less typical groups is unknown. In particular, IDDMs with onset as adults often present quite insidiously and slowly progress to insulin dependence. We were particularly interested in testing whether T cells from adult-onset type I diabetics, diagnosed after the age of 20, would evidence the IL-2 defect.

To determine this, peripheral blood lymphocytes were stimulated with either PHA, or a mixture of a phorbol ester (PdB) and a calcium ionophore (ionomycin), or anti-CD3 to induce the cells to secrete IL-2. Supernatants were isolated and tested for IL-2 content in a bioassay.

Our results indicated that these three protocols differed in how effective they were in stimulating IL-2 release. Stimulation with PdB and ionomycin induced the highest levels of IL-2 release, followed by PHA and then by anti-CD3. However, the T cells from the adult-onset IDDM patients did not differ significantly in the levels of IL-2 produced compared to age-matched controls. Presently we are investigating the levels of anti-islet cell antibodies to determine whether there is evidence for humoral autoimmunity in these individuals.

Our results suggest that defective T-cell function as defined by decreased IL-2 production may not be involved in adult-onset IDDM. This implies that the mechanism underlying disease onset in the population we tested differs from that operative in the classical juvenile-onset IDDM. A better understanding of the events leading to disease onset is important in order to investigate the possibility of such preventive therapies as immunosuppression.

**36. A 64KD Antigen Associated with Graves' Orbitopathy Is Amphiphilic and Present in Retro-Orbital Muscle, Skeletal Muscle, and Thyroid Tissue.** DL Kendler, GK Huber, J Rootman,* TF Davies, Division of Endocrinology, Mount Sinai School of Medicine, New York, NY and Department of Ophthalmopathy, University of British Columbia, Vancouver, British Columbia, Canada.

Techniques for the isolation of transmembrane (amphiphilic) proteins now permit a more detailed analysis of their antigenicity. We have, therefore, explored the recently described 64kd antigen associated with autoantibodies in the serum of patients with severe Graves' orbitopathy.

Membranes (100,000 × g) from fresh human fetal eye muscle (EM) and skeletal muscle (SM) were solubilized with 2% Triton X-114, the detergent phase separated and the extraction repeated. Denaturing and reducing SDS PAGE on a 7.5% discontinuous gel was performed using 2.5 μg of membrane antigen. Sera from 11 patients with recent onset Graves' disease and enlarged extraocular muscles on CT scanning were used for Western blotting and bound IgG detected by $^{125}$I-Protein A and autoradiography.

Silver staining of SDS-PAGE gels revealed selective partitioning of proteins between the detergent and aqueous phases. In the EM detergent phase, 64kd immunoreactivity was easily detected in all of the patients with orbitopathy ($n = 11$) and only detectable in low quantity in normal sera ($n = 4$). Reactivity to a detergent-phase SM antigen of the same size was observed with 10 of the 11 EM detergent-phase positive sera. Six sera were immunoblotted onto a detergent-phase extract of human thyroid membranes and showed interaction with a similar molecular weight antigen.

We conclude that IgG antibodies binding to SM and EM membrane 64kd antigen are present in sera from patients with Graves' orbitopathy. Furthermore, "specificity-crossover" with a similar molecular weight membrane antigen present in human thyroid tissue could be demonstrated. This 64kd antigen partitions with the detergent phase and hence may be amphiphilic or transmembrane in nature. The lack of specificity for retro-orbital muscle shown by Western blotting of patient sera questions the role of such an antigen in disease etiology and may represent a secondary immune response to muscle and/or thyroid damage.

**37. Human Thyroid Peroxidase (TPO) Expression in a Transfected Lymphoblast Line (RAJI-TPO): A System for Thyroid Autoantigen Processing and Presentation.** A Martin, RP Magnusson, V Brennan,* DL Kendler, TF Davies, Departments of Medicine and Pharmacology, Mount Sinai School of Medicine, New York, NY 10029.

Antigen-presenting cells (APCs), such as monocytes and B cells, play a critical role in the processing and presentation of antigenic fragments, complexed with MHC molecules, to antigen-specific helper T cells.

To create an effective in vitro model of thyroid autoantigen processing and expression for the study of human thyroid autoimmune disease we have transfected an immortal lymphoblastoid cell line (Raji, EBNA+, HLA-DR+) with a full-length TPO cDNA in a construct containing the $\text{Hyg}^R$ gene and the cis-acting fragment of EBV (oriP) (1).

More than 50% of RAJI-TPO cells expressed surface TPO, as judged by immunoreactivity with human polyclonal anti-TPO and flow cytometry. Western blotting revealed characteristic protein bands at 107 kDa with similar sizes to that observed in control Chinese hamster ovary (CHO) cells expressing TPO (2). Northern blotting demonstrated a 3.7 kb mRNA in both RAJI-TPO and CHO-TPO; larger than natural human TPO (3.2 kb).

The virtually unlimited growth potential of human lymphoblastoid cell lines causes RAJI-TPO to be a potentially useful resource for the study of autoantigen processing, MHC/autoantigen interaction, and presentation to the immune system.

References

1. Sugden B, Marsh K, Yates J. Mol Cell Biol 1985; 5:410–413.
2. Kendler DL, Martin A, Magnusson RP, Davies TF. Clin Endocrinol 1990; 33:751–760.

---

# Gastroenterology

**38. Effect of Hydroxychloroquine in the Treatment of Active Ulcerative Colitis: Results of the Open-Label Phase of the Controlled Trial.** L Mayer, P Turtel, D Present, H Janowitz, R Rosenberg, S Meyers, D Sachar, M Conn, S Itzkowitz, S Martin, Divisions of Clinical Immunology and Gastroenterology, Box 1089, Mount Sinai Medical Center, NY 10029, (212) 241-7351.

Hydroxychloroquine (HQ) has been used in the treatment of a number of autoimmune disorders. While its efficacy has been proven, its mechanism of action is unclear. We have described a defect in intestinal epithelial cells in IBD which appears to be corrected by HQ in vitro. We recently completed a double-blind placebo-controlled trial testing the efficacy of HQ (400 mg/day) in 90 patients with steroid-dependent (ST) and nonsteroid-dependent (NS) active ulcerative colitis.

Following the 6-week blinded portion of the trial, the patients had the option of being placed on open-label drug for a period of 6 months.

Of the 67 patients with at least 6 weeks of follow-up on medication, 42 demonstrated clinical and/or sigmoidoscopic improvement (62.6%), 22 achieving clinical remission (32.8%), including 3 in whom remission was induced during the trial. Of those patients who improved, most had moderate disease activity at initial assessment: four patients who achieved clinical remission flared after cessation of steroids, 3 requiring colectomy. Of the 25 patients who showed no improvement, 21 were stable and 4 worsened (3 during steroid taper). A total of 26 patients completed at least 6 months of HQ; 18 (69%) improved, 11 (42%) achieving remission. 52 patients had sigmoidoscopic follow-up, 33 demonstrating improvement and 5 achieving normal appearance. Within the steroid-treated (ST) group ($n = 36$), 27 improved (75%), 14 achieving remission (39%). A total of 30 ST patients (83%) were able to taper their steroids (14 off and five <5 mg); 6 could not be tapered.

Adverse reactions requiring cessation of the drug occurred in 3 patients (ocular toxicity [color blindness] in 2 patients after 12 and 16 months of therapy). The ocular toxicity was minor in both cases. Thus HQ appears to be a safe, effective, and steroid-sparing agent in the treatment of mild to moderate ulcerative colitis.

**39. Effect of a Cancer-Associated Mucin Antigen on Natural Killer (NK) Cell Function.** S Ogata, S Itzkowitz, GI Research Lab and Dept.

of Medicine, Box 1069, The Mount Sinai Hospital, New York, NY 10029, (212) 241-6749. Proc Am Assoc Cancer Res 1991; 32:234.

40. **Sialosyl-Tn Antigen (S-Tn) in Nondysplastic Biopsies May Predict Development of Cancer or Dysplasia in Ulcerative Colitis (UC).** A Kornbluth, D Present, P Rubin, DB Sachar, I Gelernt, A Greenstein, A Chen, SH Itzkowitz, GI Research Lab & Depts of Medicine and Surgery, Box 1069, Mount Sinai School of Medicine, New York, NY 10029, (212) 241-6749. Gastroenterology 1991; 100:A221.

41. **Sialosyl-Tn: A Novel Mucin Antigen Associated with Poor Prognosis in Gastric Cancer Patients.** JL Werther, S Rivera-MacMurray, H Bruckner, SH Itzkowitz, GI Research Laboratory and Depts. of Medicine and Neoplastic Disease, Box 1069, The Mount Sinai Hospital, New York, NY 10029, (212) 241-6749. Proc Am Assoc Cancer Res 1991; 32:241.

42. **Expression of a Cancer-Associated Mucin Antigen in Crohn's Colitis.** A Roston, N Harpaz, L Oberman, A Greenstein, SH Itzkowitz, GI Research Lab and Departments of Medicine, Pathology, and Surgery, Box 1069, Mount Sinai School of Medicine, New York, NY 10029, (212) 241-6749. Gastroenterology 1991; 100:A611.

43. **6-Mercaptopurine Is Effective in Intractable Proctosigmoiditis.** MA Loev, PH Rubin, ML Chapman, DH Present, Division of Gastroenterology, Department of Medicine, Mount Sinai School of Medicine, New York, NY 10029, (212) 427-4398. Gastroenterology 1991; 100:A832.

44. **6-Mercaptopurine (6-MP) in Inflammatory Bowel Disease (IBD) Patients of Childbearing Age: No Increase in Congenital Anomalies—a Case Controlled Study.** A Dayan, P Rubin, M Chapman, D Present, Department of Medicine, Mount Sinai School of Medicine, New York, NY 10029. Gastroenterology 1991; 100:A824.

45. **Serum Lipase: A Better Test to Diagnose Acute Alcoholic Pancreatitis.** V Gumaste, P Dave, G Sereny, Dept. of Medicine, City Hospital Center at Elmhurst, New York, NY 11373, (718) 830-1752.

The utilization of serum amylase levels to diagnose acute pancreatitis in alcoholic patients with abdominal pain has certain drawbacks: serum amylase may be elevated in alcoholics without acute pancreatitis, and some alcoholics with pancreatitis may have normal serum amylase levels.

Therefore, we undertook a prospective study to determine whether the estimation of serum lipase would be a better test in these circumstances. Serum lipase was estimated in 202 asymptomatic alcoholics without abdominal pain (group A) and the levels were compared with those found in 26 patients with image-proven (ultrasound or CT scan) acute pancreatitis of alcoholic etiology (group P). The values (in U/L) were:

| | S. Amylase | | S. Lipase | |
|---|---|---|---|---|
| | Range | Mean ± SD | Range | Mean ± SD |
| Control | 40–100 | 62 ± 17 | 31–213 | 94 ± 55 |
| A | 17–347 | 71 ± 36 | 34–600 | 186 ± 111 |
| P | 180–2985 | 714 ± 685 | 1011–25,706 | 5371 ± 5432 |

Serum amylase: Values up to thrice normal were seen in asymptomatic alcoholics. Twelve of the 26 patients with pancreatitis had levels that overlapped those found in asymptomatic alcoholics. The sensitivity of a serum amylase level >3 normal in detecting pancreatitis was 54%. Specificity was 100%. Serum lipase: serum lipase levels up to thrice normal were also found in group A. However, none of the 26 patients with pancreatitis had serum lipase levels that overlapped with those found in asymptomatic alcoholics, and a level >3 normal had 100% sensitivity in picking out patients with acute pancreatitis. Specificity was 100%.

Conclusion: Serum lipase is a better test than serum amylase to diagnose acute episodes of alcoholic pancreatitis.

# Geriatrics

46. **Vasopressin Hypersecretion in Transgenic Mice.** Myron Miller, Shigeki Kawabata, Michelle Wiltshire-Clement, Jon W Gordon, Division of Geriatric Medicine, Departments of Medicine and Geriatrics and Adult Development, Box 1070, Mount Sinai School of Medicine, New York, NY 10029, (212) 241-4057.

Alterations in vasopressin (AVP) secretion and action occur as part of the normal aging process. In order to develop animal models of physiologic changes with aging, we have created transgenic mice carrying the rat gene for vasopressin.

The rat AVP genomic sequence with 5kb of 5′ flanking material was microinjected into the pronucleus of fertilized mouse ova which were subsequently implanted into pseudopregnant CDI female mice. Two male and one female offspring were identified by Southern blotting to carry the rat gene. Lines of mice heterozygous and homozygous for the rat AVP gene have been generated by breeding of the male carriers with normal B6D2F1 females. Studies have been carried out in one line (F-6) to characterize the function of the transgene. Measurement of AVP mRNA by RNAse protection assay has demonstrated expression of both rat and mouse message in the hypothalamus. Endogenous mouse mRNA was partially suppressed in the heterozygous hypothalamus and almost completely suppressed in the hypothalamus of the homozygous animal.

The hypothalami of mice homozygous and heterozygous for the rat AVP gene contained 7.0 ± 0.4 and 6.2 ± 1.1 mU AVP respectively vs 3.3 ± 0.5 mU AVP in the control animals ($p < 0.001$; $<0.05$). Neural lobe AVP content was not increased in the transgenic animals (209 ± 20 vs 193 ± 24 mU). However, frontal-temporal cortex brain tissue demonstrated increased amounts of AVP in both homozygous and heterozygous animals. Assessment of AVP secretion showed a significant increase in plasma AVP concentration of homozygous mice (27.7 ± 7.2 µU/mL) as compared to control animals (8.0 ± 2.1 µU/mL; $P < 0.03$). Both heterozygous and homozygous mice had a significant increase in 24-hour urinary AVP excretion. Twenty-four-hour urine collections showed no differences in urine volume and osmolality between the homozygotes, heterozygotes, and normal littermates.

These data indicate that mice lines which carry the rat gene for AVP are capable of expressing the gene in the hypothalamus. There may be autoregulation of endogenous AVP production as a consequence of expression of the foreign gene so that total AVP secretion is only moderately altered. The animals provide a model for study of long-term effects of vasopressin hypersecretion on AVP regulation and action.

47. **Elevated Tumor Necrosis Factor Levels in Alzheimer's Disease.** H Fillit, W Ding, J Kalman, L Altstiel, B Lawlor, G Wolf-Klein,* Departments of Geriatrics, Medicine, and Psychiatry, Box 1070, Mount Sinai Medical Center, New York, NY, (212) 241-4274, and Parker Jewish Geriatric Institute, New Hyde Park, NY.

Recent investigations have clearly demonstrated a local inflammatory response surrounding and invading the senile plaques of Alzheimer's disease (AD), including microglia and cytokines such as IL-1. Tumor necrosis factor alpha (TNF) is an important cytokine with serious potential systemic toxic effects that has been correlated with disease activity in some studies.

Levels of TNF in sera from three groups of patients with AD and age-matched controls were initially measured by an enzyme-linked immunoassay. Subsequently, a cytotoxicity bioassay was employed using L929 cells.

Significantly elevated levels of TNF were found in AD sera compared to controls in all three populations studied and in both assays.

Elevated circulating TNF may be derived from the local CNS inflammatory reaction in AD, and may account for some systemic manifestations of AD such as weight loss. Future studies may determine if, in the absence of complicating disorders which may elevate TNF, circulating TNF could be a marker of AD inflammatory activity. These studies provide further evidence of an inflammatory component to the disease process in AD.

48. **Racial Differences in Postmenopausal Bone Homeostasis.** DE Meier, MM Luckey, S Wallenstein, R Lapinski, Division of Geriatrics,

Departments of Medicine, Geriatrics, Obstetrics and Gynecology, and Biomathematical Sciences, Box 1070, Mount Sinai School of Medicine, New York, NY 10029, (212) 241-5561.

Lower fracture risk in blacks is of unknown etiology, but may be associated with slower rates of bone turnover and bone loss after menopause.

To evaluate this possibility, 81 black and 86 white, normal weight, healthy peri- and post-menopausal women were studied. Cross-sectional measurements included weight, radial (RBMD), and lumbar (L2-4BMD) bone mineral density by single and dual photon absorptiometry, calcium intake, fasting (F-u-Ca) and 24-hour (CA-24) urinary calcium excretion, fasting hydroxyproline excretion (f-u-OHPro), and serum osteocalcin (BGP) and alkaline phosphatase (AlkP).

RBMD declined significantly with years since last menstrual period (YRSLMP) in both whites (11.4%/decade or .078 gm/cm per decade, $p = .0001$) and blacks (5.1%/decade or .036 gm/cm per decade, $p = .007$). By regression analysis the decline in RBMD was linear and fit was not improved by including polynomial terms. The rate of decline in RBMD in blacks was significantly less ($p < .05$) than that in whites. Adjustment for weight improved the racial differences in slope ($p < .02$).

L2-4BMD declined significantly and linearly with years since menopause in whites (8.5%/decade or .094 gm/cm2 per decade, $p = .003$) and blacks (8%/decade or .096 gm/cm2 per decade, $p = .002$). Fit was improved by either the log transformation of 1+ YRSLMP or by adding the quadratic term in YRSLMP. The decline of L2-4BMD was not significantly different between the races either before or after adjustment for weight.

Racial differences in the markers of bone turnover in postmenopausal women were:

| | Black | White | |
|---|---|---|---|
| | mean ± SE | mean ± SE | $p$ |
| Ca intake (mg/d) | 781 ± 65 | 1205 ± 77 | .0001 |
| f-u-Ca (mg/100mL GFR) | .064 ± .006 | .092 ± .008 | .02* |
| CA-24 (mg/24 hr) | 111 ± 11 | 148 ± 10 | .02* |
| f-u-OHPro (mg/gCr) | 25.8 ± 2.9 | 34.2 ± 2.7 | .03 |
| BGP (pmol/mL)** | .43 ± .04 | .61 ± .04 | .002 |
| Alk P (IU/dL) | 77.9 ± 3.7 | 77.4 ± 2.0 | NS |

* Wilcoxon rank sum test
** measured in 33 blacks and 37 whites
NS = $p > .10$

F-u-Ca, CA-24, and f-u-OHPro excretions and BGP are significantly lower in blacks than in whites. The racial differences observed in fasting and 24-hour calcium excretion persist despite adjustment for calcium intake, which was higher in whites, or for creatinine and sodium excretions, which did not differ between the races.

Conclusions: Cross-sectionally measured RBMD, but not L2-4BMD, appears to decline at a slower rate in peri- and post-menopausal blacks than in whites. Measures of bone resorption (f-u-Ca, f-u-OHPro, CA-24) and bone formation (BGP) are significantly lower in postmenopausal blacks than in whites, suggesting that less active bone resorption and lower rates of bone loss may be, in part, responsible for the higher bone mass and lower fracture risk of blacks.

**49. Aging Effects on Basal and Exercise-Stimulated Atrial Natriuretic Hormone and Vasopressin in Healthy Humans.** Gabriel Gold, Myron Miller, Michelle Wiltshire-Clement, Division of Geriatric Medicine, Departments of Medicine and Geriatrics and Adult Development, Box 1070, Mount Sinai School of Medicine, New York, NY 10029, (212) 241-4057.

Forty healthy male and female subjects aged 22 to 64 years old were studied to determine the influence of age on circulating levels of atrial natriuretic hormone (ANH) and vasopressin (AVP) in humans both under basal conditions and after stimulation of ANH release by controlled exercise.

Supine exercise using a bicycle ergometer to 80% of maximum predicted heart rate resulted in marked increases in ANH (27.9 ± 12.8 pg/mL to 90.1 ± 54.6 pg/mL) amd AVP (0.68 ± 0.51 µU/mL to 1.55 ± 1.24 µU/mL). For ANH, subjects over 50 years old had higher baseline levels (36 ± 14 pg/mL) and response to exercise (125 ± 73 pg/mL) compared to subjects under 50 (baseline: 25 ± 11.5 pg/mL; postex: 81 ± 45 pg/mL). There was no effect of age on basal or exercise-stimulated AVP levels. Comparison of ANH and AVP responses to exercise revealed an inverse correlation between peak changes in ANH and post-exercise levels of AVP.

Thus, ANH appears to have a physiologic role in modifying AVP response to exercise. These data demonstrate that exercise is a potent stimulus for both ANH and AVP release in older as well young individuals. Our results provide further support for an enhancing effect of aging on ANH release. Aging does not appear to affect a nonosmotic stimulus for AVP release.

**50. Roadblocks to Do-Not-Resuscitate Orders: A Study in Policy Implementation.** BE Cammer Paris, VC Carrion, JS Meditch Jr, CF Capello, Department of Geriatrics & Adult Development, Box 1070, Mount Sinai School of Medicine, New York, NY 10029, (212) 241-5561.

This study examines obstacles encountered by physicians in the implementation of a do-not-resuscitate (DNR) order on a hospitalized patient.

Sixty house staff officers and 45 attending physicians at two New York City medical centers responded to a questionnaire that listed 18 potential problems encountered in obtaining a DNR order. Using a Likert scale of 1 to 5 (1 = rare; 5 = common), that data was rank-ordered, based on the total number of points scored for each question from all respondents (the highest-ranked question was the one which received the highest number of points out of a potential maximum of 525 points) (see Table 1). Additional data were collected on characteristics of the physicians, including prior instruction in procedure for obtaining DNR orders, age, gender, and extent of postgraduate training (see Table 2).

Analysis of the data indicates that the major obstacles to obtaining a DNR order are the paucity of discussions between physicians and patients regarding DNR issues, as well as situations involving surrogate decision-making. The problems most commonly encountered involving surrogates include inability to contact a surrogate, surrogate disagreement with a DNR order, disagreement between potential surrogates about a patient's DNR status, and uncertainty as to who is the appropriate surrogate. Other problems, including the filling out of cumbersome forms, are not identified as important obstacles in DNR implementation.

This study suggests the need for improved communication between physicians, patients, and surrogates about these issues in advance of hospitalization.

Table 1: Rank Order of Responses by Point Value

| Rank | Question | Total Points | Condensed Statement |
|---|---|---|---|
| 1 | 3 | 385 | AttP does not discuss DNR with pt |
| 2 | 12 | 381 | Surrogate not contacted before pt arrests |
| 3 | 10 | 357 | Surrogate does not agree to DNR |
| 4 | 14 | 341 | Disagreement between potential surrogates about pt's DNR status |
| 5 | 13 | 338 | Uncertainty about who is proper surrogate |
| 6 | 7 | 303 | Communication failure: AttP & HS |
| 7 | 1 | 301 | Pt refuses DNR status |
| 8 | 8 | 278 | AttP fails to write order |
| 9 | 6 | 274 | Disagreement: AttP & HS about pt's DNR status |
| 10 | 5 | 253 | HS not comfortable discussing DNR with pt |
| 11 | 11 | 248 | Surrogate agrees verbally, does not sign DNR |
| 12 | 15 | 218 | Surrogate objects to DNR order; pt previously agreed |
| 13 | 4 | 217 | HS too busy to discuss DNR with pt |
| 14 | 16 | 204 | Disagreement AttP & HS about pt's capacity |
| 15 | 2 | 191 | Pt agrees verbally but does not sign DNR |
| 16 | 9 | 186 | HS too busy to complete DNR forms |
| 17 | 17 | 140 | Witness unavailable |
| 18 | 18 | 133 | Witness refuses to sign DNR |

AttP = attending physician; HS = house staff; pt = Patient.

Table 2: Rank Order of Responses by Point Value

| | Beth Israel | Mount Sinai | Totals |
|---|---|---|---|
| Total Responses | 55 | 50 | 105 |
| Attending physicians | 25 (45%) | 20 (40%) | 45 (43%) |
| Age | | | |
| Median | 45 | 45 | 45 |
| Range | 29–76 | 32–66 | 29–76 |
| Gender* | | | |
| Male | 24 (96%) | 16 (80%) | 40 (89%) |
| Female | 1 (4%) | 3 (15%) | 4 (9%) |
| | | 1 NR (5%) | 1 NR (2%) |
| Years in Practice* | | | |
| Median Years | 15 | 14 | 14 |
| Range Years | 1–47 | 1–35 | 1–47 |
| Prior DNR Instruction | 9 (36%) | 4 (20%) | 13 (29%) |
| House Staff Officers | 30 (55%) | 30 (60%) | 60 (57%) |
| Age | | | |
| Median | 26 | 28 | 28 |
| Range | 23–36 | 25–33 | 23–36 |
| | 1 NR | | 1 NR |
| Gender* | | | |
| Male | 20 (67%) | 20 (67%) | 40 (67%) |
| Female | 9 (30%) | 9 (30%) | 18 (30%) |
| | 1 NR (3%) | 1 NR (3%) | 2 NR (3%) |
| Fellows | 2 | 2 | 4 |
| PGY-3 | 4 | 14 | 18 |
| PGY-2 | 9 | 4 | 13 |
| PGY-1 | 15 | 10 | 25 |
| Prior DNR Instruction | 15 (50%) | 11 (37%) | 26 (43%) |

* Not all respondents answered these questions
NR—No Response

**51. Evaluation of the Substituted Judgment Standard: How Accurate Are Proxy Predictions?** AB Seckler, DE Meier, M Mulvihill, BC Paris, Department of Geriatrics & Adult Development, Box 1070, Mount Sinai School of Medicine, New York, NY 10029, (212) 241-5561. Ann Int Med 1991 (in press).

**52. Immunochemical Studies of Autoantibodies to Vascular Heparan Sulfate Proteoglycan: Antiprotein Core Antibodies in Autoimmune Disease and the Aged.** J Pace, C MacDonald,* N Boyle, M Mulvihill, M Blake,* L Kerr, H Spiera, L Libow, J Zabriskie,* H Fillit, Departments of Geriatrics and Medicine, Box 1070, Mount Sinai Medical Center, (212) 241-4274, and The Rockefeller University, New York, NY.

Vascular heparan sulfate proteoglycan (vHSPG) is an important structural and functional molecule of the vascular barrier. Human anti-heparan sulfate (vHS) autoantibodies (autoAbs) described in a variety of clinical states have unclear immunospecificity and pathologic significance. In contrast, our laboratory has reported autoAbs reactive with purified intact vHSPG in diseases possibly involving antivascular autoimmunity including glomerulonephritis and dementia. It is not known if human autoAbs to intact vHSPG recognize vHS and/or vHSPG protein core epitopes. No previous studies have shown human autoAbs to vHSPG protein core.

Intact purified vHSPG was prepared from bovine glomeruli and vHSPG protein core was prepared by trifluoromethanesulfonic acid digestion. Characterization by SDS-PAGE and immunoblotting employing monoclonal antibodies to vHSPG protein core revealed purified vHSPG protein core at MW ~100 kd. The presence of autoAbs to vHSPG protein core in sera from patients with autoimmune disease (poststreptococcal glomerulonephritis [PSGN] and systemic lupus erythematosus [SLE]) and from aged nursing home (NH) patients was investigated by immunoblotting.

These studies clearly demonstrated the presence of IgG autoAbs to vHSPG protein core in PSGN, SLE, and aged NH patients. By ELISA, the prevalence of IgG autoAbs to purified intact vHSPG and to vHSPG protein core in NH sera and age-matched controls was investigated. Sera from approximately 30% of NH patients contained IgG autoAbs to intact vHSPG. Of the 18 aged sera reactive with intact vHSPG, 8 also reacted with vHSPG protein core by ELISA. Control sera were unreactive to intact and vHSPG protein core.

This first demonstration of IgG anti-vHSPG protein core autoAbs in disease sera suggests that autoimmunity to specific high-affinity vHSPG protein epitopes may play a role in human vascular disease. Further studies are under way in our laboratory to investigate the epidemiology and immunospecificity of vHS and vHSPG protein core autoimmunity in autoimmune disease and aging.

**53. The Role of Anthropometric Assessment for Malnutrition in the Hospitalized Frail Elderly.** S Lansey, H Fillit, Geriatric Evaluation and Treatment Unit, Departments of Geriatrics and Medicine, Box 1070, Mount Sinai Medical Center, New York, NY 10029, (212) 241-4274. Gerontologist 1990; 30:326A.

# Hematology

**54. Combined Surgery and Chemotherapy for the Treatment of Gastrointestinal Non-Hodgkin's Lymphoma.** KM Troy, J Cuttner, Division of Hematology, Department of Medicine, Box 1079, Mount Sinai School of Medicine, New York, NY 10029, (212) 241-6481.

The gastrointestinal (GI) tract is a common site of extranodal involvement in non-Hodgkin's lymphomas (NHLs). The optimal management of patients with GI involvement by NHL is controversial.

Twenty-two patients with NHL and GI involvement at diagnosis were seen by the authors at Mount Sinai Medical Center over an 8-year period. There were 11 men and 11 women with a median age of 61 years. Eleven patients (50%) had gastric involvement and seven (32%) had small bowel involvement. One patient each had involvement of the esophagus alone and of the colon alone. One patient had diffuse involvement of the ileum and colon, and one patient had discrete lesions involving the esophagus, stomach, and ileocecal region. Of the 22 patients, 21 had lymphomas of intermediate or high-grade histology; 13 had stage IE or IIE disease and 9 had stage IV disease.

Of the 13 patients with stage IE or IIE disease, 11 patients were treated with a combination of surgical resection and an adriamycin-containing chemotherapy regimen. All of these patients remain in remission 35 to 124 months following diagnosis. Of the 2 remaining patients, one had an esophageal lesion which was not surgically resectable and the other underwent gastrectomy for a gastric lesion. Neither patient received an adriamycin-containing regimen and both have relapsed 29 and 57 months after diagnosis, respectively. Of the 9 patients with stage IV disease, only 2 patients underwent attempted surgical resection. All were treated with adriamycin-containing combination chemotherapy regimens. Only 1 patient has achieved long-term disease-free survival. The majority of stage IV patients have died due to failure to control lymphomatous disease outside of the abdomen.

Complications of therapy included 4 episodes of bowel perforation. All of these occurred in stage IV patients, one of whom had undergone a partial resection of disease. All were managed surgically and none were fatal.

These results strongly suggest that localized gastrointestinal lymphoma is a curable disease when treated by a combination of surgical resection and adriamycin-containing combination chemotherapy.

**55. Lipoprotein (a) Is a Substrate for Factor XIIIa And Tissue Transglutaminase.** W Borth, PC Harpel, Division of Hematology, Department of Medicine, Box 1079, Mount Sinai School of Medicine, New York, NY 10029, (212) 241-0483.

The mechanisms which mediate deposition of lipoprotein (a) [Lp(a)], an atherogenic lipoprotein particle, onto the vessel wall and cell surfaces are unknown. An irreversible deposition of Lp(a) may require the presence of enzymes that catalyze its binding to surface-oriented structures. Transglutaminases catalyze crosslinking of proteins as well as incorporation of primary amines into protein substrates.

We studied whether tissue transglutaminase and/or activated Factor XIII (plasma derived or recombinant F XIIIa) incorporate primary amines into Lp(a).

In the presence of Ca$^{2+}$, F XIIIa and tissue transglutaminase catalyze incorporation of dansylcadaverine or [$^{14}$C] putrescine into purified Lp(a) in a specific and time-dependent manner. SDS PAGE demonstrated that DSC became incorporated into the apo (a) portion of Lp(a). Lp(a) purified from five different donors showing different apo (a) phenotypes were substrates for TGs. Western blot analysis confirmed that apo (a) was the major dansylcadaverine carrying protein moiety of Lp(a). Tissue TG also extensively crosslinked the apo (a) portion of the Lp(a) particle. Characterization of the specificity of tissue TG showed that fibronectin, $\alpha_2$-plasmin inhibitor and apo (a) could be readily labeled with DSC by tissue TG, but other proteins including low density lipoprotein [LDL], IgG, $\alpha_1$-proteinase inhibitor and albumin showed poor or no reactivity. Direct comparison of Lp(a) with LDL showed that apo B 100 was a poor substrate for transglutaminases. Recombinant apolipoprotein (a) proved to be an excellent substrate for TGs in that one mole recombinant apolipoprotein (a) incorporated as much as 15 moles of [$^{14}$C]putrescine, which corresponded to five times the amount of amine incorporated into Lp(a).

The susceptibility of Lp(a) to transglutaminases suggests a mechanism whereby the interaction of Lp(a) with surface receptors and other surface-oriented structures could be enzymatically altered.

# Hypertension

**56. A Model for Evaluating Ambulatory Blood Pressure Monitoring in Management of Mild Hypertension.** Clyde Schechter, Marianne Fahs, Lawrence R. Krakoff, Mount Sinai Medical Center, New York, NY 10029.

Usual screening methods for hypertension may be insufficiently specific in selecting for drug treatment. Ambulatory blood pressure monitoring has potential for bias-free and more accurate determination of average pressure.

We have derived computer-generated data from a simulated population of 10,000 subjects by Monte Carlo methods from bivariate distributions of casual and ambulatory pressures developed from Framingham and other published studies. Two strategies were chosen to select for drug treatment. The 1988 Joint National Committee recommendations (JNCR) led to selection of 23%. JNCR followed by ambulatory monitoring and use of echocardiography (for borderline cases) led to 6%.

Despite the differences in percentage allocated to treatment, there was no difference in calculated mean survival, 38.2 ± 0.4 (SE) yr vs 38.0 ± 0.4 yr.

The predictions of this model have important implications for a cost effective approach to management of mild hypertension.

# Infectious Diseases

**57. Infections in Liver Transplant Patients: The First 100 Patients.** BR Meyers, AC Gurtman, MH Mendelson, SZ Hirschman, W Dewell, M Schwartz, C Miller, Division of Infectious Diseases, Departments of Medicine and Surgery, Box 1090, Mount Sinai School of Medicine, New York, NY 10029, (212) 241-6741.

To evaluate the infectious complications of liver transplantations, the transplant records of 100 patients, mean age 48 years, who underwent 124 orthotopic liver transplants at The Mount Sinai Hospital in the 2-year period 9/88 to 8/90 were reviewed.

Pretransplant diagnoses included hepatitis C, 27; hepatitis B, 3; cirrhosis: alcoholic 16, primary biliary 13, cryptogenic 7 and sclerosing cholangitis 11. There were 149 episodes of infection (1.5/patient) and 96 episodes of rejection. Sites of infection included liver hepatitis, 41; pneumonia, 21; wound infection, 20; septicemia, 19; peritonitis, 4. Of 151 organisms, the most common were cytomegalovirus (CMV) 29, methicillin-resistant *S. aureus* (MRSA) 16, *Enterobacter* sp. 14, *Candida* sp. 11, *Pseudomonas* sp. 11, hepatitis C 10, hepatitis B 1, herpes simplex 8, and *Aspergillus* sp. 3. The overall mortality was 21%: 11 (52%) infection, 4 (19%) bleeding, 3 (14%) liver failure, and 2 (9%) rejection. No temporal relationship between rejection and infection was noted. Increase in infection with OKT3 usage was noted. Liver transplant complications: GI bleeding 16, seizures 13, acute renal failure 11, hepatic artery thrombosis 6, portal vein thrombosis 2, and bile leak 7. Twenty-nine episodes (25 patients) had CMV infection post op; 22 (76%) hepatitis, 3 (10%) pneumonia, 4 (14%) GI; prior recipient CMV serology: 63 positive, 30 negative; 10+/63+ (15%), 12−/30− (40%) developed infection with CMV. Twelve negative recipients received 6+/6− donor (Do); 10 positive recipients received 4+/6− donor.

The most common cause of infection was hepatitis and CMV; death was most commonly due to infection.

**58. *Pseudomonas aeruginosa* Bacteremia in AIDS Patients.** MH Mendelson, S Szabo, AC Gurtman, EP Neibart, M Policar, BR Meyers, GS Hammer, S Reddy, SZ Hirschman, Division of Infectious Diseases, Department of Medicine, Box 1090, Mount Sinai School of Medicine, New York, NY 10029, (212) 241-6741.

Objective: To determine the clinical features, risk factors, relapse rate, and outcome of *P. aeruginosa* (PSA) bacteremia in AIDS patients (P).

Methods: Eighteen episodes (E) of PSA bacteremia (12 primary (1°), 6 relapse) in 12 AIDS P were evaluated for demographic features, acquisition site (community-acquired, CA; nosocomial, N), source, risk factors, laboratory parameters, therapy and outcome.

Results: Of the 1° E, 8 were CA; 4 were N. There were 11 males, 1 female (× age 40); 6 IVDUs, 6 homosexuals; diagnosis of AIDS was carried for 1–48 (× 20) mos. Sources of bacteremia included: 8 pneumonia (cavitary 5), 2 sinusitis, 2 otitis (1 malignant, MEO), 5 indwelling catheter associated, 2 unknown. Five P had a past medical history (PMH) of sinusitis (4) or otitis (2). Radiographic features of the 8 E of pneumonia included: 3 bilateral, 5 cavitary and 1 focal infiltrate(s). WBCs ranged from .1–26.2 (× 4.1) × $10^3$/mm$^3$, 11/15 absolute neutrophil counts were >1 × $10^3$/mm$^3$ in 18 E. 1° N E occurred from 13–30 (×20) days after admission. Anti-PSA therapy (1 agent) was instituted at presentation in 6/12 1° E; all relapses received 2 agents. Mortality secondary to evaluable E of PSA bacteremia was 47% (33% 1°, 67% relapse); occurred within 72 hrs and was not associated with institution of anti-PSA therapy at onset.

Conclusions: PSA is often an unanticipated etiologic agent of CA and N bacteremia in AIDS P. Characteristic features include predominance of upper and lower respiratory tract sources (72%), PMH of sinusitis/otitis (42%), cavitary pneumonia (63% of pneumonias), and high relapse rate (42%) and mortality (47%).

**59. Epidemiologic Study and Policy Analysis of a Measles Screening/Vaccination Program in an Urban Medical Center.** MH Mendelson, N Fallick, JT Dong, D Rose, E Bottone, D Hodes, L Finkelstein, BR Meyers, G Haskins, SZ Hirschman, Division of Infectious Diseases, Departments of Medicine, Pediatrics, Nursing, and Administration, Box 1090, Mount Sinai School of Medicine, New York, NY 10029, (212) 241-6741.

Objective: As a result of the upsurge of measles throughout the U.S., hospitals have implemented vaccination programs for health care workers (HCWs). In order to determine the cost-effectiveness of measles (M) immunity (I) testing (T) prior to vaccination (VA) (Screen First, SF) versus VA without screening (Vaccinate ALL, VALL) an analysis of a SF program in 8/90 was conducted.

Methods: 328 HCWs who did not meet CDC criteria for MI (born before 1957; receipt of 2 doses live measles vaccine, LMV; or physician-dx M) were tested.

Results: 39 (11.9%) were (−) and subsequently given 2 doses LMV without complications. SF cost incurred/100 HCWs (supplies, labor, EIA MIT at $3.20, LMV 2 doses at $14.82) was $1218. If all HCWs were given 2 doses of LMV without MIT (VALL), estimated cost incurred/100 HCWs is $2802. SF is < VALL by $1584/100 HCWs. SF costs/100 HCWs would have to exceed $2465 to justify a VALL program.

Conclusions: Under optimistic assumptions (susceptibles 5%, MIT $5, LMV 2 doses $50), SF results in a savings of $4230/100. Under pessimistic assumptions, (susceptibles 30%, MIT $40, LMV 2 doses $20), SF costs more. This model is applicable to hospitals choosing strategies for measles vaccination programs.

**60. Intramuscular (IM) Pentamidine for *Pneumocystis carinii* Pneumonia (PCP).** TW Cheung, R Matta, E Neibart, G Hammer, E Chusid, HS Sacks, Division of Infectious Diseases and Clinical Trials, Department of Medicine, Box 1090, Mount Sinai School of Medicine, New York, NY 10029, (212) 241-6741.

Objective: To evaluate retrospectively the efficacy and toxicities of IM pentamidine for PCP prophylaxis.

Methods: Patient eligibility: history of PCP or HIV infection with CD4 count less than 210/mm$^3$, sulfa intolerance, baseline neutropenia, or intolerant or unsuitable for aerosolized pentamidine. Monthly IM pentamidine was given by Z-track technique at a dosage of 300 mg or 4 mg/kg if weight was less than 50 kg.

Results: Eighty-nine patients have been treated by this method. Their age ranged from 22 to 57 (median 35). Risk groups were 31 homosexual men, 37 intravenous drug users, 14 heterosexual partners of HIV infected men, 2 blood transfusion acquired, and 5 unknown risk factor. Twenty-eight patients had ARC, 54 had confirmed AIDS (34 with prior history of PCP) and 7 patients were asymptomatic. Duration of therapy was from 1 to 28 months (median 8 months; 706 total patient-months). Three patients developed PCP. Adverse reactions were two episodes of hypotension, three cases of sterile abscess, and three patients developed glucose intolerance.

Conclusion: The use of IM pentamidine for PCP prophylaxis appears to be effective and has minimal toxicities. A prospective clinical trial is necessary to confirm this result and to compare efficacy with other methods.

**61. Influenza and Pneumococcal Vaccination for HIV-Infected Patients: A Policy Analysis.** DN Rose, CB Schechter, HS Sacks, Divisions of General Medicine and Infectious Disease, Department of Medicine and Department of Community Medicine, Mount Sinai School of Medicine, New York, NY 10029. Dr. Rose: Box 1009, (212) 241-1897.

Objective: To analyze the policy of vaccinating human immunodeficiency virus (HIV)-infected patients against influenza and pneumococcal infections.

Methods: We performed a decision analysis on hypothetical cohorts of HIV-infected patients of various degrees of immunodeficiency (no, moderate, and severe immunodeficiency). Monte Carlo simulation and published data were used to calculate when cohort members move from one immune state to another. A Markov model was used to simulate the risks and benefits of vaccination over 10 years. AIDS experts were surveyed for estimates of the unknown variables: risk for HIV-infected patients relative to non-HIV-infected patients of severe infections (requiring hospitalization), case fatality rates, and vaccine effectiveness.

Results: Though vaccine effectiveness diminishes slightly with advanced HIV disease, the risks of pneumococcal infection rise substantially. Pneumococcal vaccination was therefore found to be a reasonable prevention strategy at all HIV disease stages: 71–224 hospitalizations and 47–126 deaths are prevented over 10 years per 100,000 cohort, depending upon the cohort's CD4 count at the beginning; 447–1401 vaccinations must be given to prevent one hospitalization and 685–1911 vaccinations must be given to prevent one death.

By contrast, influenza vaccination was found to be a less efficient policy, because case fatality is lower, immunodeficiency increases influenza risks only minimally, and vaccination is necessary annually: although 78–159 hospitalizations are prevented over 10 years per 100,000 cohort, only 10–15 deaths are prevented, depending upon CD4 count at the beginning; however, 4701–6617 vaccinations must be given to prevent one hospitalization and 51,359–55,678 vaccinations must be given to prevent one death. Limiting influenza vaccination to those with CD4 counts above 200 or 500 × $10^6$ cells/L has little effect on the ratio of vaccinations to deaths.

Conclusions: Pneumococcal vaccination is a reasonable prevention strategy for HIV-infected patients at all stages of immunodeficiency. Fewer deaths are prevented by influenza vaccine, at a cost of many more vaccinations, making influenza vaccination a far less efficient prevention strategy.

# Liver

**62. Comparative Analysis of the Histology of Hepatitis C and Autoimmune Chronic Active Hepatitis.** N Bach, SN Thung, F Schaffner, Depts of Medicine, Box 1101, (212) 369-6433 (Division of Liver Disease) and Pathology, Mount Sinai School of Medicine, New York, NY 10029.

Prior to the advent of a serologic marker for hepatitis C, morphologic features of this diagnosis could not be confirmed.

Methods: We examined biopsies from 50 patients with chronic hepatitis/cirrhosis-C (HCV Ab+ –ELISA method) (CAH-C) and 21 patients with autoimmune chronic active hepatitis (AiCAH). A blinded observer graded each biopsy based on 19 different histological features.

Results: At time of biopsy, the average age of patients with CAH-C was 46 vs 36 years for AiCAH. Etiology of CAH-C was blood transfusions, 50%; history of intravenous drug abuse, 26%; sporadic, 16%; and occupational risk, 8%. Cirrhosis was seen more frequently in AiCAH (90%) than CAH-C (58%) [$\chi^2$, $p = 0.01$], while duration of disease until biopsy was 1.7 years for AiCAH vs 14.3 for CAH-C. Features more common with CAH-C were bile duct damage (91 vs 40%) and loss (91 vs 20%), steatosis (72 vs 19%) and lymphoid follicles (49 vs 10%) [$p = 0.01$]. More severe grades of lobular necrosis (76 vs 31%), lobular inflammation (76 vs 38%) and piecemeal necrosis (81 vs 10%), as well as the presence of multinucleated giant cells (29 vs 6%) and areas of collapse (76 vs 6%) were seen in AiCAH [$p = 0.01$]. Five with AiCAH received immunosuppression prior to biopsy; exclusion of these patients accentuated differences seen. Different patterns of inflammation and necrosis were also noted.

Conclusions: Morphologic criteria, in addition to serologic data, are useful for differentiating CAH-C from AiCAH. Histologically AiCAH is a more aggressive disease than CAH-C.

**63. The Effect of Ethanol on Taurocholate Uptake by Rat Hepatocytes in Primary Monolayer Culture.** J-Z Ni, BJ Potter, Division of Liver Diseases, Department of Medicine, Box 1039, Mount Sinai School of Medicine, New York, NY 10029, (212) 241-8879.

Chronic alcoholism is associated with an increased incidence of pigment gallstones and decreased cholesterol stone formation. Although ethanol alters bile acid metabolism (and bile flow), what effect it has on uptake is not known. The aim of this study was therefore to examine the effect of ethanol on taurocholate uptake by cultured rat hepatocytes.

To test this, hepatocytes were isolated from rat livers by collagenase perfusion and cultured on Biomatrix in serum-free medium containing 100 mM ethanol for up to 7 days. [$^3$H]Taurocholate uptake was determined at a range of concentrations and $V_{max}$ and $K_m$ determined. Conditioned medium was prepared by culturing nonparenchymal cells, isolated by pronase digestion of the liver, in the presence or absence of ethanol.

In control cells, $V_{max}$ remained constant for up to 7 days (434 ± 44 vs. 419 ± 93 pmol/min per mg protein for day 1 and 7 respectively). In the presence of 100 mM ethanol, $V_{max}$ dramatically increased over the same period (858 ± 30 vs. 393 ± 63 pmol/min per mg protein on day 7, $P < 0.001$). However, the $K_m$ for both groups also increased significantly from 42 ± 6 μM to 125 ± 12 μM in controls and from 37 ± 5 μM in the presence of 100 mM ethanol. The addition of conditioned medium had no effect on control cells, but did reduce both $V_{max}$ and $K_m$ for cells cultured in 100 mM ethanol, from 609 ± 79 to 307 ± 16 pmol/min per mg and 136 ± 6.8 to 54.6 ± 4.3 μM, on day 3.

Hepatocytes were isolated from rats pretreated (5 g/kg per day) with ethanol by gavage and cultured for 24 hours. No difference was seen in the $K_m$, although the $V_{max}$ still increased (981 ± 159 pmol/min per mg vs. 450 ± 1.5 pmol/min per mg) on day 7.

These results suggest that exposure to ethanol for 3 days or more increases the uptake of taurocholate both in vivo and in vitro in the rat. The difference observed between these two models may reflect modulation of taurocholate uptake by the nonparenchymal cell in the liver.

**64. Acute and Chronic Effects of Ethanol on Ferritin Uptake by the Liver.** H Zhang, BJ Potter, Division of Liver Diseases, Department of Medicine, Box 1039, Mount Sinai School of Medicine, New York, NY 10029, (212) 241-8879.

Alcoholism is often associated with liver siderosis and an increase in serum ferritin levels. It is now known that ferritin can function as a serum iron transporter as well as an iron storage protein. Furthermore, a hepatocyte receptor-mediated uptake mechanism and a plasma membrane ferritin receptor have now been characterized for this protein. It was therefore the aim of this study to investigate the effects of ethanol on isolated rat hepatocytes to assess the contribution of serum ferritin to the siderosis of alcoholic liver disease.

This hypothesis was tested using the following methods: Rat ferritin was radiolabeled with $^{59}$Fe in vivo and isolated from rat livers by conventional techniques. Hepatocytes were isolated from rat livers of either naive rats, or animals pair-fed the Lieber-DeCarli liquid alcohol diet for 4–6 weeks, by collagenase perfusion. These cells were then incubated at 37°C with [$^{59}$Fe]ferritin for up to 3 hours.

Ferritin uptake by freshly isolated hepatocytes was the same as in cells cultured overnight, with a $V_{max}$ of 1.785 ± 0.2 ng ferritin/min per $10^6$ cells. Incubation in the presence of ethanol resulted in a progressive reduction in uptake with increasing alcohol concentration, from 1.23 ± 0.05 ng ferritin/$10^6$ cells per min (control) to 0.65 ± 0.02 ng/$10^6$ cells per min in 100 mM ethanol ($p < 0.001$). 4-Methylpyrazole 0.1 mM restored 70% of this activity ($p < 0.05$). Five μM cyanamide decreased ferritin uptake slightly in the presence of ethanol (0.82 ± 0.04 ng ferritin/$10^6$ hepatocytes per min vs 0.86 ± 0.03 ng/$10^6$ cells per min). Preincubation with acetaldehyde resulted in a dose-dependent reduction to approximately 25% with 300 μM acetaldehyde ($r = 0.858$, $p < 0.001$). Ethanol thus has an acute inhibitory effect on ferritin uptake, which is mediated by acetaldehyde production. Ferritin uptake by hepatocytes from alcohol-fed rats, however, was markedly increased (0.703 ± 0.021 ng/min per $10^6$ cells) over the pair fed controls (0.493 ± 0.028 ng/min per $10^6$ cells).

These studies suggest that the direct metabolism of alcohol by hepatocytes inhibits the ferritin uptake. However, long-term alcohol abuse appears to cause an increase of ferritin uptake and may be responsible for the hepatic siderosis seen in alcoholics.

**65. The Plasma Membrane Fatty-Acid Binding Protein and Mitochondrial Isoform of Aspartate Aminotransferase of Rat Liver Are Closely Related.** D Stump, SL Zhou, BJ Potter, D Sorrentino, LM Isola, PD Berk, Division of Liver Diseases (Departments of Medicine and Biochemistry), Box 1039, Mount Sinai School of Medicine, New York, NY 10029, (212) 241-8319.

In 1985 we reported the isolation of a fatty-acid binding protein of ca 40 kDa from Triton X-100 extracts of sinusoidally enriched rat liver plasma membranes by affinity chromatography over oleate agarose and elution with 8M urea (FABP$_{PM}$[1]). Antibodies to FABP$_{PM}$[1] inhibited long-chain fatty acid uptake in rat hepatocytes, adipocytes, and cardiac myocytes. Subsequently, to achieve higher yields, we used salt extraction and HPLC, involving successive elutions from hydroxyapatite and hydrophobic interaction columns. The resulting protein, designated FABP$_{PM}$[2], exhibited physicochemical characteristics identical to FABP$_{PM}$[1]. The two proteins gave a line of identity in Ouchterlony studies and reacted identically in Western blots with rabbit antibodies against either antigen. Both anti-FABP$_{PM}$[1] and anti-FABP$_{PM}$[2] inhibited cellular fatty acid uptake, and stained hepatocyte plasma membranes by indirect immunofluorescence. We conclude that FABP$_{PM}$[1] and FABP$_{PM}$[2] are identical.

Subsequent evidence, including molecular size, pI, distribution of multiple isoforms, aminoacid analysis, N-terminal aminoacid sequence, fatty acid affinity, and aspartate aminotransferase (AspAT) enzymatic activity suggest, unexpectedly, that FABP$_{PM}$[2] and mitochondrial AspAT (mAspAT) are closely related. The two proteins behave similarly in Western blots and cross-react in Ouchterlony studies; anti-mAspAT also inhibits cellular fatty acid uptake. In fractionated 3T3-L1 cells, Western blots employing anti-FABP$_{PM}$[2] demonstrate a 43 kDa band in mitochondrial pellets from both the fibroblast and adipocyte phases; a similar band is observed in purified plasma membrane fractions only from 3T3-L1 adipocytes, appearing coincidentally with facilitated fatty acid uptake during adipocyte differentiation.

Although these studies suggest that mAspAT and FABP$_{PM}$ are similar, RNAse protection assays on mRNA isolated from 3T3-L1 fibroblasts and adipocytes, using a probe derived from the 5′ end of an authentic mAspAT cDNA, demonstrate the appearance of a new, smaller band which hybridizes with the probe coincident with the appearance of the plasma membrane FABP and facilitated fatty acid uptake during adipocyte differentiation. Hence, while FABP$_{PM}$ and mAspAT are similar, they may not be identical.

**66. Development of a Saturable Oleic Acid Uptake Mechanism in Adipocytes upon Differentiation from 3T3-L1 Fibroblasts Is Associated with the Surface Expression of a 43 kDa Membrane Fatty-Acid Binding Protein.** SL Zhou, D Stump, D Sorrentino, BJ Potter, LM Isola, PD Berk, Division of Liver Diseases, Box 1039, Mount Sinai School of Medicine, New York, NY 10029, (212) 241-6479.

We have demonstrated that fatty-acid (FFA) uptake kinetics in rat hepatocytes, adipocytes, cardiac myocytes, and jejunal enterocytes are consistent with a facilitated transport process, and have identified a 43 kDa plasma membrane FFA binding protein (FABP$_{PM}$), antibodies to which selectively inhibit FFA uptake in these cell types. This protein shares several properties, including molecular size, pI, N-terminal aminoacid sequence and immunoreactivity, with the mitochondrial isoform of aspartate aminotransferase (mAspAT).

To further investigate the role of FABP$_{PM}$ in FFA transport and its relationship to mAspAT, we have examined the uptake of [$^3$H]-oleate (OA) by 3T3-L1 cells using a rapid filtration technique. 3T3-L1 cells, derived from mouse embryo fibroblasts, differentiate in culture to cells possessing the morphological and enzymatic characteristics of mature adipocytes. In incubations containing 500 μM albumin (BSA) the OA/BSA ratio was varied from 0.1 to 2.0, and the uptake of OA by confluent 3T3-L1 fibroblasts and differentiated adiopcytes was studied.

The uptake by adipocytes was saturable (Vmax = 110 pmol/sec per mg protein, Km = 53 nM unbound OA in solution), was consistently greater (by ≥5-fold) than the comparable uptake by fibroblasts, and was inhibited by a monospecific polyclonal rabbit anti-rat FABP$_{PM}$. This same antibody, and also an anti-mAspAT, revealed surface FABP$_{PM}$ by immunofluorescence or immunohistochemical methods on 3T3-L1 adipocytes, but not on fibroblasts. Partially purified mitochondria and purified plasma membranes (PM) were prepared from both of the cell lines. The specific activity ratios of PM marker enzymes (purified PM/crude cell homogenates) for both the fibroblast and adipocyte stages indicated a 5–7 fold enrichment of PM and a 7–12 fold deenrichment of mitochondria in the PM preparations. Western blots employing either the monospecific anti-FABP$_{PM}$ or an analogous anti-mAspAT identified a single 43 kDa band in the mitochondrial pellets from both fibroblasts and adipocytes. However, the two antibodies detected a 43 kDa protein only in PM isolated from adipocytes, but not in the same concentration of PM protein from fibroblasts. RNAase protection assays on mRNA isolated from 3T3-L1 fibroblasts and adipocytes, using a probe derived from the 5′ end of an authentic mAspAT cDNA, demonstrate the appearance of a new, smaller band which hybridizes with the probe coincident with the appearance of the plasma membrane FABP and facilitated fatty acid uptake during adipocyte differentiation.

Conclusions: Differentiation of 3T3-L1 fibroblasts into adipocytes involves development of a facilitated plasma membrane FFA uptake mechanism and simultaneous expression of a plasma membrane fatty-acid binding protein. These data support the hypothesis that FABP$_{PM}$ plays an important role in facilitated, transmembrane fatty-acid transport. Although FABP$_{PM}$ and mAspAT share numerous characteristics, the RNAase protection assay suggests they may not be identical.

## Medical Oncology

**67. An Increase in Nonviable Differentiated Leukemia Cells following Combination Cytotoxic-Differentiation Therapy for Acute Promyelocytic Leukemia.** S Waxman, S Weinrauch, Y Huang, Mount Sinai School of Medicine, Division of Medical Oncology, New York, NY 10029.

All-trans retinoic acid (ATRA) induces clinical remissions which appear to result from differentiation in previously untreated or chemotherapy-resistant acute promyelocytic leukemia (APL). Therefore, we have used HL-60 cells as a model for APL to design therapeutic protocols to enhance differentiation and cytotoxicity.

Our data demonstrates: (1) Synergistic induction of differentiation in HL-60 cells following treatment with 5-fluorodeoxyuridine (5-FUdR) with leucovorin and the differentiation inducers ATRA, DMSO or HMBA; (2) Dexamethasone acts synergistically with ATRA to induce differentiation in HL-60 cells; (3) 5-FUdR, but not 5-fluorouracil (5-FUra) cytotoxicity is enhanced by DMSO, HMBA, or ATRA in HL-60 cells; (4) HL-60 cells resistant to adriamycin could be induced to differentiate by DMSO, HMBA or ATRA; (5) The synergistic cytotoxicity that occurs following exposure to 5-FUra/HMBA in mouse erythroleukemia cells (MEL) is associated with the development of DNA double-strand breaks; (6) In contrast to MEL cells, the inhibition of mouse bone marrow colony formation by 5-FUra was not enhanced when followed by HMBA; (7) thymidine prevents the enhanced 5-FUdR and ATRA cytotoxicity and differentiation; (8) MEL cells resistant to 5-FUra would still be induced to differentiate by HMBA and continue to demonstrate the synergistic cytotoxic effect of 5-FUra/HMBA.

As a mode of treatment, combination cytotoxic-differentiation therapy with 5-FUdR and ATRA could result in an increase in nonviable differentiated leukemia cells, resulting in a more durable clinical remission without the disseminated intravascular coagulation associated with conventional cytotoxic chemotherapy programs.

# Molecular Medicine

68. **Cloning and Characterization of a New Early Serum Responsive Gene in Vascular Smooth Muscle.** Stephen D Wax, Claire-Lise Rosenfield, Mark B Taubman, Molecular Medicine Division, Department of Medicine and Brookdale Center for Molecular Biology, Box 1128A, Mount Sinai School of Medicine, New York, NY 10029, (212) 241-1766.

Abnormal growth of vascular smooth muscle cells (VSMC) is a major feature of atherosclerosis and hypertension. In order to elucidate the genes controlling the growth of VSMC, we have utilized the technique of differential screening of cDNA libraries. This procedure enables us to compare the steady-state levels of particular mRNA species present in VSMC during different growth states. Quiescence, in which the population of cultured cells as a whole is undergoing minimal growth, can be generated by maintaining the cells in 0.3% calf serum for 48 hours. Proliferation of these cells can then be induced by the addition of PDGF. We have focused on identifying genes whose expression is altered early after the stimulation of mitogenesis because we believe that these are the most likely to regulate the cell's entrance into the proliferative state.

Thus a cDNA library was constructed in Lambda ZAPII from mRNA isolated 3 hours after the addition of 20 ng/mL PDGF-BB to quiescent VSMC. From this library we have isolated a number of clones encoding PDGF-inducible genes.

Sequencing of one such clone (pRASM 20), which contains a 2.5 kb insert, has failed to reveal significant nucleotide homology with any gene in the GenBank. In addition, the derived amino-acid sequences from all putative open reading frames of pRASM fail to show significant homology to the proteins contained in the NBRF. This strongly suggests that pRASM 20 encodes a novel PDGF-inducible gene. On Northern blot analysis, pRASM 20 hybridizes to a 3.0 kb mRNA which is expressed at low levels by quiescent VSMC (baseline) and is rapidly induced by 10% calf serum or PDGF; mRNA levels peak at 1–2 hours and then drop quickly below baseline by 4 hours. pRASM 20 does not hybridize with mRNA from 3T3 fibroblasts or COS cells.

The time course of expression suggests that pRASM 20 is a new member of the family of "immediate early" genes induced by PDGF. In addition, pRASM 20 may show specificity for VSMC.

# Neoplastic Diseases

69. **Trilineage (TLR) Response of Myelodysplastic Syndromes to Subcutaneous (SQ) Azacytidine.** LR Silverman, JF Holland, D Nelson, G Clamon, BL Powell, CD Bloomfield, R Larson, EP Demakos, SL George, RB Davis, E Liu, B Davis, C Schiffer, OR McIntyre, Departments of Neoplastic Diseases and Medicine, Box 1129, Mount Sinai School of Medicine, New York, NY 10029, (212) 241-6368, and the Cancer and Leukemia Group B, Lebanon, NH 03766.

Myelodysplastic syndromes are heterogeneous disorders characterized by ineffective hematopoiesis often resulting in death from bone marrow failure or leukemia. We previously demonstrated, in CALGB 8421, the efficacy of continuous IV infusion azacytidine in patients with myelodysplastic syndromes, producing a 38% trilineage response (Proc Am Soc Clin Oncol 8:198, 1989).

In CALGB 8921, azacytidine was administered at the same total dose, 75 mg/m$^2$ per d via SQ bolus injection for 7 consecutive days repeated q28 d. Of 66 patients entered, 42 are currently evaluable for response and toxicity. Median age for all entered is 66 years (range 23–82); 45 (68%) are male and 51 (88%) had performance status 0–1. Myelodysplastic syndrome subcategories included: refractory anemia with excess blasts (RAEB), 45%; refractory anemia with excess blasts in transformation (RAEB-T), 43%; and chronic myelomonocytic leukemia (CMML), 12%. The myelodysplastic syndrome was therapy-related in 11 (17%). Median entry blood counts were: Hgb 89 g/L, WBC 3.6 × $10^9$/L, platelets 39 × $10^9$/L. Clinically significant bleeding was evident in 41% and 79% required RBC transfusions.

Of 42 evaluable patients, 26 (62%) have responded thus far, with a median follow-up of 4 months (range 1–12): 3 (7%) have achieved a CR (complete normalization of bone marrow and blood); 7 (17%) have achieved a PR (≥50% restitution of the deficit from normal for each of the red, white, and platelet cell counts); 16 (38%) are improving (≥25% and ≤50% restitution of the deficit from normal of the blood counts &/or ≥50% decrease in transfusions). Of those improving, 14% have trilineage response, but not yet PR; 17% have bilineage response; and 7% have monolineage response or loss of transfusion requirement. The major toxicity of azacytidine is nausea and vomiting in 83%: grade 0–1 in 67%, grade 2 in 24%, grade 3–4 in 9%. Trilineage response (CR and PR) has thus occurred so far, in 24% of patients treated with SQ azacytidine.

Agents with multilineage activity are essential for the majority of patients with myelodysplastic syndromes. The trilineage response to azacytidine appears superior to that reported for IL-3, G-CSF or GM-CSF. Bone marrow hypoplasia occurred in <10%. SQ azacytidine is a cost-effective, well-tolerated, ambulatory therapy for patients with myelodysplastic syndromes.

70. **Long-Term Survival of Patients Failing Cisplatin Treatments for Stage III-IV Ovarian Cancer.** HW Bruckner, CJ Cohen, R Wallach, B Kabakow, M Chesser, J Mandeli, JF Holland, Mount Sinai Medical Center, New York, NY 10029-6574.

Patients (pts) were preregistered to cisplatin regimens: P; PT; AP; CAP; CHAP I and CHAP II.

Of the pts with only microscopic residual at second-look, 9 of 18 survived 5 yrs. Pts failing primary treatment with CHAP I or II, with macroscopic residual disease at second-look, had a median survival of 24 mo, with 1/17 alive at 5 yrs. In contrast, pts failing AP-CAP or high-dose cisplatin with CHAP available as salvage therapy had a 57-mo median survival with 6/19 alive at 5 yrs (*p* .04). The introduction of CHAP as primary therapy was associated with a significant improvement of overall survival (Bruckner, AACR 1990). The availability of CHAP also allows salvage of some patients only identifiable by second-look surgery.

There is still little published information on the long-term outcome of intensive salvage efforts. This large systematic experience supports and extends the conclusion that there are some 5-yr survivors in all groups "failing" cisplatin regimens. The chances of surviving 5 yrs appears to have increased in recent years. Some of these 5-yr survivors are disease-free. Some were pts with large tumors at the time of failure, 10/144 based on clinical evidence of progression. None had borderline tumors. In deciding criteria and survival end-points for salvage chemotherapy trials, short-term response and in particular short-term survival are no longer adequate.

71. **Survival in Stage II-III Pancreatic Cancer (PC).** HW Bruckner, MR Chesser, J Mandeli, Mount Sinai Medical Center, New York, NY 10029-6574.

Effective neo-adjuvant therapy provides an investigational setting for defining both the indications and the optimum time for PC surgery.

An initial group of 20 patients (pts) with both stage II and metastatic stage III PC lived a median of 12 months (mos) following the start of RT (6500 cGy split course)-Fluorouracil 1,000 mg/m$^2$ d1-5 continuous infusion, Streptozotocin 350–500 mg/m$^2$ d1-3 with escalation, and Cisplatin 100 mg/m$^2$ d3 (FSP). RT-FSP was followed by Leucovorin (LV)-FU maintenance treatments. RT-FSP produced local control: 35% complete remission and 30% partial remission; the remainder had stable disease. Only 3 of 20 tumors grew in the RT field, all more than 4 mos after RT. Pts with large (15 were ≥5 cm) stage II and III tumors have a survival curve for the 0–24 mos period equal to that of contemporary stage I pts treated with primary complete resection at the same institution. Additional accrual brings the number to 36 stage II pts, many normally ineligible for RT because of large tumor size, including 5 pts converted to eligibility for RT-FSP with MLP-F chemotherapy. Survival is 85% at 6 mos, 50% at 12 mos, and 20% at 2 years. Classic refractory pain has been largely eliminated. Among pts referred for treatment, 80% of stage II pts and 60% of stage III pts were, by GITSG chemotherapy criteria, eligible for treatment. In the initial group, side effects, all reversible, included 1 infection, 1 nephrotoxicity, and 4 grade 4 toxicities. There were no deaths and 1 early end of treatment due to toxicity.

The new finding, excellent local control of PC, adds to the evidence that combined modality therapy is safe neo-adjuvant therapy. Combined modality treatment for inoperable pts competes with surgery for operable pts and is safer than surgery. This level of effectiveness supports investigation of neo-adjuvant treatment for PC followed by elective surgery ± intraoperative RT.

**72. Polypeptide Hormones: Three Examples of a New Frontier in Biochemical Modulation.** HW Bruckner, JM Lee, RB Erlich, T Ohnuma, Mount Sinai Medical Center, 1 Gustave L. Levy Place, New York, NY 10029-6574.

Somatostatin analogue SMS 201-995 (SMS) inhibits human breast cancer, cell line MCF-7, as measured by both cell count and thymidine incorporation assay. Seeding of MCF-7 cells on well plates, in the presence of SS for 6 days, produced inhibition of colony formation, maximal inhibition at $10^{-8}$ M in an inverted bell curve from $10^{-10}$–$10^{-6}$ M. In contrast, preincubation with SMS for 6 days followed by SMS-free reseeding resulted in prior SMS dependent ($10^{-10}$–$10^{-6}$ M) increase in colony formation from 60% to >400%. Flow cytometry studies in the presence of SMS demonstrated marked shift in DNA content to the left, suggestive of a hypodiploid cell population. That SMS increases the absolute number of colonies, and recruits or releases "stem cells," a putative virulence characteristic, warrants further study and attempts at synchronization for subsequent therapy with S phase cytotoxins.

SMS increased DOX ($10^{-6}$–$10^{-5}$ M) antitumor effect. Clonogenic MCF-7 assays followed 5-day exposure to SMS and further exposure to both SMS and either FU (24h exposure) or DOX (24h exposure) on day 6. SMS alone produced an increase in the number of colonies. Although FU produced cell kill in a dose-dependent manner, the number of colonies was consistently greater in the presence of SMS (30–40%). In contrast, adding SMS to DOX vs. DOX alone, decreased colony counts. SMS plus DOX warrants further preclinical study.

Mice that received SMS plus FU had higher nadir WBCs (FU 500:1098/μL vs. FU 500 + SMS:3086/μL; FU 400:2426/μL vs. FU 400 + SMS:4290/μL; FU 300:1650/μL vs. FU 300 + SMS:3686 μL). That SMS improves nadir blood counts extends findings of Maurer et al. (Arzneimittelforschung 1984) that SMS could inhibit BCNU granulocyte colony unit and antitumor cytotoxicity. The mouse appears to serve as a realistic clinical model both for developing protective effects and potential beneficial interactions of SMS with chemotherapy.

# Pulmonary

**73. Sleep Apneas in Multiple System Atrophy with Autonomic Failure.** A Bhuptani,* LK Brown, H Kaufmann, C Stacy, MF Sloane, Division of Pulmonary and Critical Care Medicine and Department of Neurology, Mount Sinai School of Medicine, New York, NY 10029.

Multiple system atrophy with autonomic failure (MSA-AF or Shy-Drager syndrome) is associated with degeneration of brainstem nuclei which may be involved in respiratory control. During sleep, respiration is controlled by brainstem neurons with no input from higher centers. There are reports of sudden death during sleep and sleep apneas in these patients. The purpose of this study was to evaluate a series of such patients for sleep architecture, number and types of sleep apneas, associated desaturation, and cardiac dysrhythmias.

Nine patients (5 male, 4 female) with MSA-AF were evaluated. Mean duration of illness was 5.7 years ± 3.99 SD and mean age was 63.2 years ± 13.4 SD. Two patients had tracheostomy at time of study, done for vocal cord dysfunction. Overnight polysomnography was performed using a 10 channel grass 78 D polygraph, recording EEG, EKG, EOG, oral and nasal flow, oxygen saturation (BIOX 3700 pulse oximeter), breath sounds, and chest and abdominal contribution to tidal volume (respiratory inductive plethysmography with semiquantitative calibration). Apnea was defined as 4% desaturation from baseline with no or ineffectual air movement (tidal volume <150 cc) for at least 10 seconds. Hypopnea was defined same as apnea, but less than 10 sec in duration, or air movement reduced by at least 50% from baseline with desaturation. Apnea, hypopnea, and combined apnea + hypopnea (respiratory disturbance) indices were calculated. AI, HI or RDI higher than 10 were considered abnormal.

Sleep architecture in these patients was abnormal. Seven of 9 patients had no REM sleep. Stage III-IV non-REM sleep was reduced in 7/9 (mean value, 10%, nl 20–25%) and the stage I-II non REM sleep was increased in all 9 patients (mean value, 88%, nl 45–60%). Respiratory disturbance index (RDI) and hypopnea index (HI) were abnormally high in all 9 patients; respective mean values, 47 and 38. Apnea index (AI) was abnormally increased in 4/9 patients; mean value, 27. Both patients with tracheostomy had a normal AI, although one had a few central apneas. Apneas were predominantly of central type (61%), followed by obstructive (26%) and mixed (13%). Apneas and hypopneas were associated with desaturation and respective mean minimum saturation were 84% and 88%. Baseline mean saturation was 95%. No significant cardiac dysrhythmias were noted.

In conclusion, this study showed altered sleep architecture and abnormally high RDI and HI in patients with MSA-AF. AI was abnormally high in some patients. Apneas were predominantly of central type. These episodes were associated with serious desaturation, but no cardiac dysrhythmias.

**74. Does Cotton Bract Extract Challenge Cause "Monday" Bronchospasm?** J Russomanno, J Siegel, E Fine, N Rienzi, TJ Witek, MG Buck, J Godbold, EN Schachter, Division of Pulmonary, Department of Medicine, Box 1232, Mount Sinai School of Medicine, New York, NY 10029, (212) 241-5900.

Byssinosis is an occupational airway disease associated with the inhalation of cotton dust. "Monday dyspnea" has been used to describe the associated symptoms accompanied by objectively measured bronchospasm on the first workday of the work week. Cotton bract extract (CBE) is a water-soluble extract used to characterize byssinotic responses. We studied the ability of CBE to cause "Monday" bronchospasm.

We exposed 19 healthy nonsmokers (NS) to CBE inhalation on 5 consecutive days, and on a 6th day, following a 2-day interruption of CBE inhalation (day 8). Partial and maximal expiratory flow volume loops were obtained prior to baseline (BL) and at 15-min intervals for 2 hours after CBE inhalation. Changes from BL in the maximal expiratory flow at 40% of the vital capacity on the partial curve (MEF 40% (P)) were measured, and the maximal MEF 40% (P) percent decrement from BL occurring during the 2 hours following CBE inhalation was studied for each subject on each day.

The average maximal MEF 40% (P) percent decrement from BL (AMMPDB) was calculated for each of the 6 days; the AMMPDB was 14.3 ± 3.4% on day 1, 9.2 ± 4.0% on day 2, 11.3 ± 3.8% on day 3, 8.7 ± 3.4% on day 4, 9.2 ± 3.5% on day 5, and 13.5 ± 3.6% on day 8. Longitudinal data analysis, comparing the AMMPDB from day 1 with the AMMPDB from each of the subsequent days, revealed that the AMMPDB was greater on day 1 than on day 2 ($p = .11$), day 4 ($p < .05$) and day 5 ($p = .07$). No significant difference in AMMPDB was noted between day 1 and day 6 ($p = .76$). No significant difference in AMMPDBs was noted between non-Monday days.

Conclusion: (1) Acute decrements in expiratory flow rates were observed following sequential daily challenges with CBE of cotton bract extract, (2) daily decrements in expiratory flow rate following inhalation of CBE on 5 consecutive days were greatest on the first day, (3) after a 2 day interruption of CBE inhalation, CBE induced decrements in expiratory flow rate return to levels noted on the first day of CBE inhalation.

Supported by NIOSH 5-RO1-OHO2593.

**75. Antioxidant Therapy and Survival in ARDS: A Clinical Study.** PL Marino, MJ Sawyer, Division of Pulmonary and Critical Care Medicine, Department of Medicine, Box 1232, Mount Sinai Medical Center, New York, NY 10029, and the Department of Surgery, The Graduate Hospital; University of Pennsylvania, Philadelphia PA.

The adult respiratory distress syndrome (ARDS) is an inflammatory disorder of the lungs that is untreatable and often fatal. Oxidant injury has been implicated in the pathogenesis of ARDS.

The present study evaluated antioxidant therapy in ARDS using glutathione surrogates (N-acetylcysteine and selenium) and hydroxyl radical scavengers (alpha-tocopherol and ascorbate). All subjects were adults with respiratory failure requiring mechanical ventilation, and the diagnosis of ARDS was made using a combination of chest x-ray findings, pulmonary artery occlusion pressure, and the clinical setting. Patients were assigned to receive either conventional supportive therapy (N = 20) or an antioxidant regimen (N = 25) consisting of: (1) N-acetylcysteine, 150 mg/kg po as loading dose, followed by 70 mg/kg po every 4 hrs, (2) alpha-tocopherol, 400 IU po every 6 hrs, (3) selenium, 50 mcg IV every 6 hrs, and (4) ascorbic acid, 500 mg IV every 6 hrs. Antioxidant therapy was started within 24 hours of the diagnosis of ARDS, and was continued for 72 hours. Only patients who could tolerate enteral fluids received the antioxidant regimen. Patients in each group were compared at entry using three clinical scoring systems: APACHE II (range 0–55), Organ Failure Index (Range 0–6), and Lung Injury Score (Range 0–4).

Entry characteristics and clinical outcome at eight weeks (results reported as mean ± (SD) were:

| | CONTROL | ANTIOXIDANTS |
|---|---|---|
| | N = 20 | N = 25 |
| Entry | | |
| APACHE II | 22.8 (6.39) | 26.8 (7.58)* |
| Organ failure | 3.05 (1.02) | 3.00 (0.71) |
| Lung injury | 3.18 (0.48) | 3.12 (0.86) |
| OUTCOME @ 8 wks | | |
| Weaned from ventilator | 7 (35%) | 21 (84%)** |
| Discharged from ICU | 7 (35%) | 20 (80%)** |
| Mortality | 65% | 20%** |

* $0.01 < p < 0.05$
** $P < 0.01$

The clinical scores show both groups to be comparable except for higher APACHE II scores in the antioxidant group.

Despite sicker patients by APACHE II criteria, antioxidant therapy was associated with marked improvements in resolution of respiratory failure (as judged by weaning from ventilator) and in survival. These results support the role of oxidant injury in ARDS, and this represents the first report of improved survival in ARDS since it was first described 24 years ago.

76. **Inhibitor of Endopeptidase 24.15 Potentiates Bradykinin-Induced Contraction of Guinea Pig Trachea.** EN Schachter, E Zuskin, C Cardozo, J Godbold, N Rienzi, M Lesser, Division of Pulmonary, Department of Medicine, Box 1232, Mount Sinai School of Medicine, New York, NY 10029, (212) 241-6067.

Previously it was shown that inhibition of endopeptidase 24.11 (EP24.11) intensifies bradykinin-induced contraction of guinea pig tracheal strips. Recently we have found that the lung contains another zinc-metalloendopeptidase (EP24.15) that degrades bradykinin (ARRD 141:A727, 1990).

To evaluate the possible role of this enzyme in the lung, tracheas from Hartley male guinea pigs (n = 19) were studied in an organ bath containing 20 mL of Krebs Hanseliett solution at 37°C. Four rings obtained from each animal were incubated for 15 minutes with (1) leucine thiorphan (LT) (10 μM), an inhibitor of EP 24.11; (2) N-[(1RS)-carboxy-3-phenylpropyl]-Ala-Ala-Tyr-Para-aminobenzoate (I-24.15) (10 μM), an inhibitor of EP 24.15; (3) LT (10 μM) + I-24.15 (10 μM); and (4) Krebs solution (control). Cumulative dose response contractions were induced with bradykinin (B) by serially adding B in concentrations from 1 pM to 30 μM. Concentration response curves were computer fitted to a logistic equation, and Emax (EC50) and slope were derived:

| | Emax* (%) | EC50 (M) | Slope |
|---|---|---|---|
| Control | 22.4 | $7.3 \times 10^{-10}$ | 0.38 |
| LT | 39.2 | $7.6 \times 10^{-11}$ | 0.51 |
| I-24.15 | 26.5 | $7.4 \times 10^{-10}$ | 0.46 |
| LT + I-24.15 | 55.0 | $2.7 \times 10^{-10}$ | 0.53 |

* Expressed as percent of tissue response to maximal carbacol contraction ($10^{-5}$ M).

Supported by NIOSH 5-R01 0H02593

A significant difference ($p < 0.01$) was demonstrated for Emax between I-24.15 and LT-treated animals when compared to control. Additionally, a significant difference ($p < 0.01$) was also demonstrated between LT-treated tracheas and those treated with LT + I-24.15.

We conclude that both EP 24.11 and EP 24.15 play a role in modulating Kinin-induced contractions of the airway.

# Renal

77. **Permeabilities and Ionic Conductances in Renal Failure.** RD London, MS Lipkowitz, RG Abramson, Division of Nephrology, Department of Medicine, Box 1243, Mount Sinai School of Medicine, New York, NY 10029.

Previous studies have demonstrated that cAMP modulates relative ionic chloride permeability (PCl/PK) of normal human red cell ghosts and rat renal cortical brush-border membrane vesicles (BBMV). Since elevated endogenous [cAMP] and increased PCl/PK were also observed in RBCs of untreated human uremic subjects, the present studies were performed to evaluate whether ionic permeabilities and [cAMP] are altered in the 75% nephrectomized rat model of renal failure.

Relative ionic permeabilities were determined in red cell ghosts and renal cortical (BBMV) using the potential sensitive fluorescent probe DiS-C3-(5). Since significant hypertrophy occurs in the remnant kidney following 75% nephrectomy, uninephrectomized animals were pair-fed 40% protein.

Kidney weight was significantly higher than in sham-operated controls fed 20% protein (1.90 ± 0.04 vs 1.39 ± 0.03, $p < 0.001$), but not different from 75% nephrectomized animals (1.85 ± 0.05). Despite the significant hypertrophy, there were no detectable differences in either relative ionic permeabilities of BBMV or renal cortical homogenate [cAMP] in the high-protein uninephrectomized animals compared to controls. Renal function was significantly reduced by 75% nephrectomy: BUN was increased (27.5 ± 0.4 vs 14.3 ± 1.4, $p < 0.001$) and GFR, estimated from the creatinine clearance, was substantially reduced (2.03 ± 0.16 vs 0.67 ± 0.05, $p < 0.001$). As depicted below, PCl/PK, but not PNa/PK, was significantly increased in RBC ghosts and renal cortical BBMV prepared from renal failure animals. RBC and renal cortical homogenate [cAMP] were increased in 75% nephrectomized animals.

| | Control | Remnant | P |
|---|---|---|---|
| RBC ghosts (n = 7) | | | |
| $[KCl]_{in}$ | 84.7 ± 0.27 | 77.2 ± 2.40 | <0.05 |
| PCl/PK | 0.32 ± 0.02 | 0.41 ± 0.01 | <0.01 |
| PNa/PK | 0.69 ± 0.04 | 0.76 ± 0.04 | NS |
| BBMV (n = 7) | | | |
| $[KCl]_{in}$ | 67.8 ± 1.3 | 69.2 ± 0.86 | NS |
| PCl/PK | 1.13 ± 0.09 | 1.54 ± 0.13 | <0.005 |
| PNa/PK | 0.68 ± 0.04 | 0.71 ± 0.06 | NS |
| PCl/PNa | 1.67 ± 0.11 | 2.20 ± 0.21 | <0.05 |
| PRb/PK | 1.00 ± 0.07 | 1.14 ± 0.13 | NS |
| cAMP | | | |
| plasma (pm/mL) | 25.8 ± 5.8 | 23.2 ± 8.4 | NS |
| RBC (pm/mL RBC) | 46.3 ± 9.6 | 57.3 ± 7.4 | <0.05 |
| Kidney (pm/mg prot) | 4.30 ± 0.58 | 5.25 ± 0.61 | <0.005 |

These studies demonstrate that the relative chloride permeability of RBCs and renal BBMV is selectively increased in association with elevated tissue [cAMP] in the 75% nephrectomy rat model of renal failure. Of note, these changes are qualitatively similar to those detected in RBCs of untreated uremic humans. These cumulative findings suggest that membrane ionic permeability may be altered in diverse tissues in renal failure in association with an elevation of intracellular [cAMP].

78. **Changes in Gene Expression Induced by Renal Ischemia.** S Saggi, J Megyesi, P Price, R Safirstein, Division of Renal Diseases, Departments of Medicine and Biochemistry, Box 1243, Mount Sinai School of Medicine, New York, NY 10029, (212) 241-6727.

Recovery from renal ischemia requires replacement of necrotic tubule cells to restore the continuity of the renal epithelium. During this process, normally quiescent renal cells increase their nucleic acid synthesis and undergo mitosis. A striking parallel exists between the changes in gene expression induced by ischemic renal damage and that induced by growth factor stimulation of quiescent cells in culture. Expression of c-fos and Egr 1, prototypes of nuclear binding members of the immediate early genes, increase almost immediately after release of renal hilar clamping. Their induction is short-lived and precedes the peak of renal DNA synthesis. Likewise, increased KC and JE expression also precedes the peak of renal DNA synthesis. Such changes in immediate early gene expression after growth factor stimulation are due to increased transcription rather than increased mRNA stability. Ischemia profoundly reduces prepro epidermal growth factor (EGF) and Tamm-Horsfall (TH) gene expression as well, but these reductions occur after the onset of the immediate-early gene response; the mRNA levels remain low for a prolonged period of time. The thick ascending limb and distal tubule are the only renal sites of EGF and TH synthesis, and JE and EGR 1 have been immunocytochemically localized to the thick ascending limb during ischemia.

These studies suggest that the thick ascending limb and distal tubule remain transcriptionally active after ischemia in vivo and demonstrate both positive and negative changes in gene expression. Further studies examining the transcriptional activity of the thick ascending limb of EGF, TH, KC, and JE during experimental ischemia should provide important clues to their functional significance and lead to novel strategies designed to maximize the regenerative potential of the kidney.

---

# Rheumatology

79. **Scleroderma in Association with the Use of Bleomycin: A Report of Three Cases.** Leslie D Kerr, Harry Spiera, Division of Rheumatology, Department of Medicine, Box 1244, Mount Sinai School of Medicine, New York, NY 10029, (212) 241-6792.

Although an association between exposure to bleomycin and the development of scleroderma has been suspected, few cases are reported.

We are reporting the development of cutaneous scleroderma in three patients coincident with the use of bleomycin in low cumulative doses of less than 100 U.

Dramatic clinical improvement occurred in two patients treated with high dose steroids.

This series of three patients supports a causal connection between bleomycin and scleroderma. In light of the widespread use of bleomycin, this complication may be underappreciated.

---

# Spinal Cord Damage Research Center

80. **Gastrointestinal Transit after Spinal Cord Injury: Effect of Cisapride.** SK Rajendran, JR Reiser, MA Korsten, RL Zhang, WA Bauman, Spinal Cord Damage Research Center, Spinal Cord Injury and Medical Services, 130 West Kingsbridge Road, Bronx VA Medical Center, Bronx, NY 10468, (212) 584-9000, Mount Sinai School of Medicine (CUNY), NY 10029.

Heartburn, bloating, and abdominal discomfort are common problems in spinal cord injury (SCI) patients but, despite its clinical significance, little is known about the gastrointestinal effects of spinal cord injury. In humans, gastric dilatation and ileus occur after traumatic cord transection, which suggests the presence of abnormal gastrointestinal motility.

To study this issue, gastric emptying and mouth-to-cecal transit time (MCTT) were measured in quadriplegics and paraplegics. We also studied the effect of Cisapride, a new prokinetic agent, on these parameters. The effect of Cisapride on gastric emptying and MCTT were studied in 6 quadriplegics (C4-C6), 3 paraplegics (T12-L2), and 6 healthy controls in a randomized, double-blind, placebo-controlled crossover study. The mean age of patients was 43 years (range 19–70 years) and that of controls 39 years (27–71 years). The mean duration of injury was 10.3 years (range 7 months–33 years). Cisapride was administered orally (10 mg four times a day) for four days. Gastric emptying of liquids was studied using 200 $\mu$Ci of (Technetium-99) DTPA in 200 mL of normal saline and of solids using 300 $\mu$Ci of (Technetium-99) sulfur colloid in a scrambled egg white preparation. None of the subjects were receiving drugs known to alter gastrointestinal transit. The MCTT was measured using the hydrogen breath test (Quintron 12i) after ingestion of 10 gm lactulose in 100 ml of water.

Mean MCTT (127 ± 48.2 min, mean ± SD) in quadriplegic subjects was significantly prolonged compared to that in normal controls (55 ± 12.2 min) ($p < 0.01$). Administration of Cisapride resulted in normalization of the MCTT (62 ± 19.4 min). Transit time in paraplegics was not significantly different from that in normal controls (72.3 min). Gastric emptying of solids and liquids was normal in quadriplegics and was not changed by Cisapride.

In conclusion, MCTT is prolonged in quadriplegics, an abnormality corrected by Cisapride administration. In contrast, paraplegics had normal gastrointestinal transit. Unlike healthy subjects with an intact nerve supply to the gut, Cisapride did not appear to alter gastric emptying in SCI patients. Changes in gastrointestinal transit after spinal cord injury and the effect of Cisapride may, in part, be explained by an imbalance between parasympathetic and sympathetic outflows to the gastrointestinal tract.

---

# Oral Presentations

1. **Developmental and Tissue-Specific Regulation of the Aortic Smooth Muscle Inositol 1,4,5-trisphosphate Receptor Gene.** A Infosino, C Rosendorf, M Mendlowitz, MG Taubman, AR Marks, Divisions of Molecular Medicine and Cardiology, Department of Medicine and Brookdale Center for Molecular Biology, Box 1126, Mount Sinai School of Medicine, New York, NY 10029, (212) 241-0309.

Regulation of intracellular calcium concentration ($[\mathrm{Ca}]_i$) controls contraction and relaxation of vascular smooth muscle. Inositol 1,4,5-trisphosphate (IP3), generated by phospholipase-C (PLC) mediated cleavage of phosphoinositides, stimulates release of $[\mathrm{Ca}]_i$ by activating the IP3 receptor (IP3R), a large intracellular calcium-release channel (CRC) on the endoplasmic reticulum. The IP3R has been purified and cloned from mouse brain and purified from bovine aortic smooth muscle. In order to elucidate the structure and regulation of expression of the IP3R and the enzyme involved in generation of IP34, PLC, we have undertaken to clone cDNAs encoding both molecules.

We screened a rat aortic smooth muscle (RASM) cDNA library using cDNA probes encoding portions of the mouse aortic smooth muscle IP3R (Marks et al., JBC 1990, 265:20719) and isolated cDNA clones encoding the RASM-IP3R. This same library was screened with oligonucleotide probes encoding the rat brain PLC-α, and a full-length RASM PLC-α clone was isolated.

The RASM-IP3R is highly homologous to the mouse brain IP3R. Unlike the skeletal muscle CRC (regulated by FGF), the RASM-IP3R is constitutively expressed during differentiation in the myogenic cell line BC3H1. However, the RASM-IP3R and PLC-α are developmentally regulated in rat aortic smooth muscle. At 4 weeks after birth, RASM-IP3R and PLC-α mRNAs (by Northern blot analysis) are expressed at low levels and increase >5-fold by 12 weeks (N = 4). In contrast, in rat brain, IP3R-RASM and PLC-α mRNAs are present at equal levels throughout development from 4 to 16 weeks (N = 4).

We conclude that different mechanisms regulate the expression of smooth and skeletal muscle CRCs, and that the RASM-IP3R and PLC-α are coordinately and developmentally regulated in a tissue-specific manner.

2. **Hirudin Blocks the Effects of α-Thrombin on Gene Induction and Protein Synthesis in Vascular Smooth Muscle Cells.** DB Altmann, M Poon, JJ Badimon, V Fuster, MB Taubman, Divisions of Molecular Medicine and Cardiology, Department of Medicine and Brookdale Center for Molecular Biology, Box 73, Mount Sinai School of Medicine, New York, NY 10029, (212) 241-1766.

α-Thrombin generated during clotting may play a direct role in the response of vascular smooth muscle cells (VSMC) to arterial injury.

α-Thrombin stimulates hypertrophy of cultured adult rat aortic VSMC and activates intracellular signals associated with growth.

We report that in VSMC, α-thrombin induces the early growth response genes KC and c-myc. KC encodes a secreted neutrophil chemoattractant and is induced by balloon aortic injury in rats. C-myc is a nuclear protooncogene. Quiescent cells expressed low levels of KC and c-myc as assessed by Northern blot analysis. There was a 10-fold induction of KC mRNA after 30 min of α-thrombin treatment, with return to baseline levels by 2 hrs. A 5-fold induction of c-myc mRNA was seen after 1 hr with a more gradual return to baseline.

Hirudin, an α-thrombin inhibitor, is known to block α-thrombin-catalyzed coagulation and platelet activation, but its effect on the VSMC growth responses to α-thrombin has been less well studied. We found that hirudin completely blocked the induction of KC and c-myc by α-thrombin, but did not affect the induction of these genes by 10% calf serum. Twenty-four hour treatment with α-thrombin (40 nM) produced an 80% increase in new protein synthesis as assessed by $^3$H leucine incorporation into TCA precipitable material. Hirudin blocked the α-thrombin effect in a dose-dependent manner but did not affect the induction of protein synthesis by calf serum.

Thus, α-thrombin stimulates multiple classes of early growth response genes in VSMC. In particular, α-thrombin may directly induce the secretion of chemoattractant agents by VSMC that might play a role in the vascular response to injury. The anti-α-thrombin effects of hirudin extend beyond thrombosis and include inhibition of gene induction and protein synthesis in VSMC.

3. **Polyunsaturated Lecithin Prevents Acetaldehyde-Mediated Hepatic Collagen Accumulation by Stimulating Collagenase Activity in Lipocytes.** J-J Li, C-I Kim, MA Leo, KM Mak, M Rojkind,* CS Lieber, Section of Liver Disease & Nutrition, Alcohol Research & Treatment Ctr., 130 West Kingsbridge Road, Bronx VA Medical Ctr, Bronx, NY 10468, (212) 579-1646, Mount Sinai School of Medicine (CUNY), NY 10029, and *Albert Einstein College of Medicine, 1300 Morris Park Avenue, New York, NY 10461. Gastroenterology 1991; 100:766.

4. **A New Animal Model of Thyroid Autoimmunity.** A Martin,* H Kimura, S Thung, P Unger, P Fong, L Shultz, TF Davies, Departments of Medicine and Pathology, Mount Sinai School of Medicine, New York, NY, and *The Jackson Laboratory, Bar Harbor, ME.

To create an animal model for human thyroid autoimmune disease we have explored the severe combined immunodeficient (SCID) mouse as a recipient of human thyroid and immune cell transplants.

Following mouse thyroid ablation with radioiodine, SCID mice (n = 15) were transplanted, subcutaneously, with 3 × 3 mm thyroid tissue blocks from surgical patients with Graves' disease. All 15 mice showed significant serum levels of human thyroglobulin (range 4–253 ng/mL) 5–15 weeks after transplantation. Histologically, all thyroid transplants showed active, regular-sized follicles and were well vascularized (range 10–313 follicles/high power field). Autologous patient peripheral blood mononuclear cells ($18 \times 10^6$) were subsequently transplanted intraperitoneally into thyroid engrafted mice. The human thyroid histology after lymphocyte transplantation ranged from marked mononuclear cell infiltration and fibrosis to no change from control mice. Focal or diffuse thyroid epithelial HLA-DR expression was detected in several transplanted mice. Intravenous injection of fluorescent dye-labelled (PKH26) human lymphocytes to study thyroid-specific homing indicated the potential for lymphocyte accumulation within the thyroid transplants, although this appeared not to be organ-specific.

These studies indicate that the thyroid transplanted SCID mouse offers a particularly suitable model for the in vivo study of lymphocyte-thyrocyte interaction in human autoimmune thyroid disease.

5. **Intestinal Epithelial Cell Activation of T Cells Results in a Distinct Pattern of Tyrosine Kinase Activity.** Y Li, A Panja, L Mayer, Division of Clinical Immunology, Box 1089, Mount Sinai Medical Center, New York, NY 10029, (212) 241-7351.

The interaction of normal intestinal epithelial cells with peripheral blood T cells results in the proliferation of CD8$^+$ T cells. This event appears to be regulated by cell surface molecules on both the T cell and epithelial cell. We have previously reported that antibodies to CD8 inhibit this response, suggesting that the CD8 molecule itself may be involved in this process. Recent studies have identified a src-like tyrosine kinase on the intracytoplasmic tail of both CD4 and CD8. Activation of this kinase may result in the phosphorylation of the T-cell antigen receptor and initiate T-cell proliferation.

In order to address whether epithelial cells might operate in this manner, we stimulated peripheral blood T cells with normal epithelial cells for 5′, 30′, and 120′ in mixed lymphocyte cultures. At each time point, T cells were lysed and subjected to immunoprecipitation with mAb to phosphotyrosine. A phosphorylation reaction was performed on each precipitate. Either crosslinking of CD4 (with anti-CD4 mAb and goat anti-mouse IgG) or stimulation with conventional accessory cells (monocytes/dendritic cells) resulted in the phosphorylation of p56lck, MAP2 kinase (42kd), as well as two other bands at 116kd and 66kd. Similarly, phosphorylation of a p56 tyrosine kinase was seen when epithelial cells were used to stimulate T cells in MLR cultures. However, novel bands were detected at 97kd and 48-50kd, suggesting that an alternate activation pathway had been utilized. When T cells were cross-linked with anti-CD8 mAbs, the 48-50kd but not the 97kd band was seen.

Therefore the 48-50kd phosphotyrosine protein may reflect the cross-linking of CD8 molecules by a novel CD8 ligand on normal epithelial cells. Since epithelial cells from IBD tissue fail to stimulate CD8+ T cells in MLR, we assessed whether these cells could activate similar or different tyrosine kinases. The pattern of phosphotyrosine containing proteins was similar to that seen with conventional accessory cells, except that the epithelial cell specific 97kd band was detectable.

Taken together, these data suggest that an alternate ligand for CD8 may exist on normal intestinal epithelial cells which, when crosslinked, may activate a tyrosine kinase resulting in T cell proliferation. In IBD the failure to activate CD8+ T cells may reflect an alteration in this novel ligand. This system may represent a true physiologic role for the intracytoplasmic src-like tyrosine kinase p56lck on CD8 molecules.

6. **Molecular Cloning of the Human Cardiac Calcium Release Channel cDNA: Expression Studies in End-Stage Human Heart Failure.** AB Brillantes, AR Marks, Divisions of Molecular Medicine and Cardiology, Department of Medicine and Brookdale Center for Molecular Biology, Box 1126, Mount Sinai School of Medicine, New York, NY 10029, (212) 241-0309.

The calcium release channel (CRC) of cardiac sarcoplasmic reticulum (SR) is a large (2.3 million MW) structure required for excitation-contraction (E-C) coupling in cardiac muscle. Activation of the CRC during cardiac muscle contraction is coupled to the slow calcium channel of the transverse tubule (T-tubule) which serves as a voltage sensor. The mechanism of this coupling (E-C coupling) remains unknown. In addition, the role of the CRC in pathological conditions affecting human heart muscle (for example, cardiomyopathies) is not understood. To address these questions and to characterize the structure and function of the human cardiac CRC, we undertook to clone its cDNA.

Using rabbit skeletal muscle (SM) CRC cDNA probes, we screened a human cardiac cDNA library at low stringency and isolated cDNA encoding the cardiac CRC. Multiple overlapping cDNAs encode a 16 kb mRNA with ~70% amino acid homology to the human SM CRC. The carboxy terminal portion of the cDNA encodes hydrophobic amino acid regions which could form the calcium release channel in the SR. The remaining portion of the cDNA encodes the "foot structure" which spans the junction between the SR and the T-tubule. Northern blot analysis identifies a 16 kb mRNA in heart and brain.

To assess whether regulation of the expression of the CRC may play a role in various forms of human cardiomyopathies we examined mRNA levels in hearts from 34 patients who had undergone cardiac transplantation at the Brigham and Women's Hospital (Boston, MA) and compared these to 7 normals. Northern blot analysis was used to assess mRNA levels quantitated by densitometry (normalized to glyceraldehyde-3-phosphate dehydrogenase mRNA levels). Compared to normals, patients with ischemic cardiomyopathy (ICM, N = 15) showed a 47% decrease in CRC mRNA levels ($P < 0.005$), and patients with idiopathic dilated cardiomyopathy (N = 19) a non-significant 17% increase.

Thus, we have cloned the human cardiac CRC cDNA and demonstrated that it is downregulated in a subset of patients with end-stage ICM. Regulation of the CRC may play a role in the pathophysiology of some forms of end-stage heart failure.

# Instructions for Authors

THE MOUNT SINAI JOURNAL OF MEDICINE, established in 1934 as *The Journal of the Mount Sinai Hospital,* is a peer-reviewed general medical journal. It is indexed or abstracted in *Index Medicus, Hospital Literature Index, International Nursing Index, Excerpta Medica, Chemical Abstracts, Biological Abstracts,* and is available on major databases. Original articles, theoretical explorations, basic science articles, reviews, reports of clinical or research experience, articles on any subject affecting the practice of medicine, short notes in any specialty, letters, and articles by medical students are welcome. The *Journal* is a participant in the Agreement on Uniform Requirements for Manuscripts Submitted to Biomedical Journals.

**Communications.** Address correspondence on manuscripts review to: Managing Editor, *The Mount Sinai Journal of Medicine,* Box 1094, 19 E. 98th Street, 4D, New York, NY 10029-6574. Phone: (212) 241-6108, (212) 722-6386.

**Copyright.** The publisher reserves copyright and renewal rights on all published material, and such material may not be published elsewhere without written permission of the publisher. When submitting a manuscript, include the following assignment of copyright in your cover letter: **In consideration of The Mount Sinai Journal of Medicine (MSJM) taking action and reviewing and editing my (our) submission, the author(s) undersigned hereby transfer(s), assign(s), or otherwise convey(s) all copyright ownership to The Mount Sinai Journal of Medicine in the event that such work is published by MSJM.** Each author must sign the statement.

**Manuscript Preparation.** Submit manuscripts in triplicate, including three sets of illustrations. Manuscripts must be typewritten, **double-spaced throughout (including references, legends, and tables)**, on one side of 8½ × 11-in. white bond paper, with generous margins. Number pages consecutively. Begin a new page for title page, abstract, first page of text, references, each table, and legends. **Computer-generated typescripts** must be double-spaced and must be easy for editors and typesetters to read. **Computer tapes or disks:** consult managing editor.

**Title Page.** On the title page, on separate lines (double-spaced), include (1) title of article, phrased as concisely as possible; (2) name of each author, including first name and degree; (3) name and address of departments and institutions from which the work originated; (4) acknowledgments of sources of support; (5) name, place, and date of any conferences at which paper has been delivered; (6) the statement **Address reprint requests to . . . ,** giving full name and address, with zip code, of the appropriate author.

**Abstract.** Original articles should include an abstract, which should be a single paragraph not exceeding 150 words. State the purposes; basic approach or methods; main findings; principal conclusions. Emphasize new and important aspects. When an abstract is provided, a summary at the end is not necessary. In literature reviews and other articles which do not begin with an abstract, the manuscript should conclude with a short summary.

**Key Words.** Supply 3 to 10 key words or phrases at the bottom of the title page, for use in indexing the article. Use Medical Subject Headings from *Index Medicus.*

**Writing Style and Abbreviations.** Authors should aim for conciseness and clarity without repetition. Avoid medical jargon, abbreviated phrasing, and obscure or newly coined abbreviations. Define all abbreviations on first use, except those for standard, universally recognized units of measurement and statistical terms. Spell out such terms as white blood cell, hematocrit, superior vena cava.

**Units of Measurement.** Use SI units, giving range or interval of normal values; more familiar units may be used in addition.

**Editing.** All papers will be edited to enhance conciseness and clarity without altering meaning and to insure conformity with journal style. The author is responsible for all statements in the article.

**Illustrations.** Photographs, line drawings, graphs, and charts should increase understanding of the text. Line drawings, graphs, and charts should be professionally prepared. All illustrations should be submitted in triplicate as sharp, high-contrast glossy prints. Photomicrographs must be 5⅛ in. wide. On the back of each print affix a gum label with the number of the figure (numbered consecutively in arabic numerals as cited in text), title of manuscript, name of senior author, and the word "top" with an arrow indicating top edge. Illustrations which are to appear together should be mounted accordingly as plates and should correspond to each other in size. Protect the prints by enclosing them in heavy cardboard; do not use paper clips. Photographs of patients can be published only if a copy of the permission is submitted with the photograph. Eliminate names of patients and hospital numbers from x-rays.

**Legends.** Legends should not duplicate the text. Begin legends on a separate sheet of paper, double spaced. Each legend should be numbered to match the illustration it accompanies. Illustrations mounted as plates to appear together should be accompanied by a single legend identifying the separate elements by position ("upper left," and so on) or by letters. If letters are used in the legend, a corresponding letter must appear on the print itself. Identify all abbreviations used in a figure. Indicate magnification and stain for photomicrographs.

**Tables.** Each table should be typed on a separate page, double spaced. Number tables consecutively as they appear in text, using arabic numerals ("Table 1"). Give each table a brief title (for example, "Effect of DMSO on Rats") and type it on a separate line. If abbreviations or symbols are used in the table, identify them in a footnote.

**References.** When citing references in your text, the first work cited is numbered 1, and succeeding references are numbered in sequence; enclose the citation number in parentheses and place it before any punctuation: **Cytomegalovirus (1), steroids (2), and recreational drugs (3) have all been implicated (4).** The reference list includes only works cited in the text, numbered in the order in which they are cited. Type, double spaced, following the general arrangement and punctuation style in the examples below (journal title abbreviations conform to *Index Medicus* style):

1. Nadelman RB, Wormser GP. **A clinical approach to Lyme disease. Mt Sinai J Med 1990;57:144–156.**

2. **International Committee of Medical Journal Editors. Uniform requirements for manuscripts submitted to biomedical journals.** N Engl J **Med 1991;324:424–428.**

**When a manuscript is accepted for publication, you will be asked to confirm the accuracy of all references.**

**Proofs.** Proofs are sent to the corresponding author from the printer. Stylistic changes which do not alter meaning should not be made at this stage. Because of increased printing charges, the cost of excessive author's alterations beyond routine correction of typographical errors and major errors in data may be billed to the author. Proofs should be corrected and returned to the editorial office within 48 hours.

**Reprint Orders.** A reprint order form is sent to the corresponding author with proofs; return it to the editorial office.

THE
MOUNT SINAI JOURNAL
OF MEDICINE

Volume 58
Number 4
September 1991

## CONTENTS *continued*

GENERAL ARTICLES

Short-Term and Long-Term Possibilities for Interventive Gerontology *Gregory M. Fahy* .......... 328

STUDENT CORNER

The Changing Controversy over Surgical Resection Margins for Stage I Cutaneous Melanoma *Michele S. Green* .......... 341

ABSTRACTS

Research Day in Medicine 1991 .......... 347

LEVY LIBRARY, MT. SINAI MEDICAL CENTER

### Index to Advertisers

Eli Lilly and Company cover 2

THE
MOUNT SINAI JOURNAL
OF MEDICINE

VOLUME 58 NUMBER 5 OCTOBER 1991

# Issues in Medical Ethics
## 1. Surrogate Decision Making
## 2. Drug Use: License vs. Prohibition

Daniel A. Moros and Rosamond Rhodes, guest editors

From the Mount Sinai School of Medicine

*plus*

Edmund D. Pellegrino on curing medicine from within—the physician's personal ethics

Avi Barbasch, Richard Gorlin, Daniel Callahan, and Bernard Baumrin on bioethical challenges to physicians

Sidney M. Wolfe on the social component of medical ethics

Veterans Administration Medical Center
Bronx, New York

BETH
ISRAEL
MEDICAL
CENTER

# THE MOUNT SINAI JOURNAL OF MEDICINE

The Mount Sinai Journal of Medicine is published by The Mount Sinai Medical Center of New York and has the following affiliates: Beth Israel Medical Center, New York; Bronx Veterans Administration Medical Center, New York; and Elmhurst Hospital Center, New York.

**Editor-in-Chief**

Sherman Kupfer, M.D.

**Editors Emeriti**

David A. Dreiling, M.D.
Lester R. Tuchman, M.D.

**Associate Editors**

Harriet S. Gilbert, M.D.
Julius Wolf, M.D.

**Managing Editor**

Claire Sotnick

**Assistant Editors**

David H. Bechhofer, Ph.D.
Constanin A. Bona, M.D., Ph.D.
Edward J. Bottone, Ph.D.
Jurgen Brosius, Ph.D.
Steven T. Brower, M.D.
Lewis Burrows, M.D.
Joseph S. Eisenman, Ph.D.
Richard A. Frieden, M.D.
Steven Fruchtman, M.D.
Paul L. Gilbert, M.D.
James H. Godbold, Ph.D.
Richard S. Haber, M.D.
Noam Harpaz, M.D.
Dennis P. Healy, Ph.D.
Barry W. Jaffin, M.D.
Andrew S. Kaplan, D.D.S.
Samuel Kenan, M.D.
Suzanne Carter Kramer, M.Sc.
Mark G. Lebwohl, M.D.
Kenneth Lieberman, M.D.
Charles Lockwood, M.D.
Bernard Mehl, D.P.S.
Myron Miller, M.D.
Edward Raab, M.D.
Allan Reed, M.D.
Allan E. Rubenstein, M.D.
David B. Sachar, M.D.
Henry Sacks, M.D.
Robert Safirstein, M.D.
Ira Sanders, M.D.
Martin H. Savitz, M.D.
Clyde B. Schechter, M.D.
Michael Serby, M.D.
Phyllis Shaw, Ph.D.
George Silvay, M.D.
Barry D. Stimmel, M.D.
Nelson Stone, M.D.
Max Sung, M.D.
Rein Tideiksaar, Ph.D.
Richard P. Wedeen, M.D.
Michael F. Wesson, M.D.

**Editorial Board**

Barry Freedman, M.B.A.
Richard Gorlin, M.D.
Nathan Kase, M.D.
Panayotis G. Katsoyannis, Ph.D.
Charles K. McSherry, M.D.
Jack G. Rabinowitz, M.D.
John W. Rowe, M.D.
Alan L. Schiller, M.D.
Alan L. Silver, M.D.
Alvin S. Teirstein, M.D.

*The Mount Sinai Journal of Medicine* (ISSN No. 0027-2507; USPS 284-860) is published 6 times a year in January, March, May, September, October, and November in one indexed volume by the Committee on Medical Education and Publications (Owner), The Mount Sinai Hospital, Box 1094, One Gustave L. Levy Place, New York, NY 10029-6574. Subscription price: individuals, U.S. and Canada, $60 per year; all libraries, $65; individuals outside the U.S. and Canada, $65. Single copies, $15. New subscriptions begin with the first issue of the calendar year. Send notice of change of address 60 days before the change is effective. Second-class postage paid at New York, NY and at additional mailing offices. POSTMASTER: Send address changes to The Mount Sinai Journal of Medicine, Box 1094, One Gustave L. Levy Place, New York, NY 10029-6574. Copyright 1991 The Mount Sinai Hospital.

THE
MOUNT SINAI JOURNAL
OF MEDICINE

Volume 58
Number 5
October 1991

## CONTENTS

### ISSUES IN MEDICAL ETHICS

Daniel A. Moros and Rosamond Rhodes, editors

**Introduction** *Daniel A. Moros and Rosamond Rhodes* ............... 367

### 1. SURROGATE DECISION MAKING

### THE LAW

**Staking Out the Law** *Hon. Judith Kaye* ............... 369

**AIDS and Surrogate Decision Making** *David A. Hansell* ............... 375

**Surrogate Decisions in New York State on Sustaining Life** *Fenella Rouse* ............... 379

### THE PHYSICIAN'S EXPERIENCE: CASES AND CONFLICTS

**Infants and Children** *Ian R. Holzman* ............... 382

**Patients in Intensive Care** *Andrew B. Leibowitz* ............... 384

**Elderly Patients** *Diane E. Meier* ............... 385

### MORAL PERSPECTIVES

**Surrogate Decision Making for Incompetent Adults: An Ethical Framework** *Dan W. Brock* ............... 388

**Foster Parents as Surrogates for Infants and Young Children** *Jonathan D. Moreno* ............... 393

### PANEL DISCUSSION

**The Ethics of Surrogate Decision Making** *Moderators: Daniel A. Moros and Rosamond Rhodes* ............... 398

### 2. DRUG USE: LICENSE VS. PROHIBITION

### SOCIAL AND SCIENTIFIC BACKGROUND

**History of Drug Use and Drug Control** *Daniel M. Fox* ............... 403

**The Epidemiology of Cocaine Use in New York State** *Blanche Frank, Don C. Des Jarlais, Rozanne Marel, James Schmeidler, and Michael Maranda* ............... 406

*continued*

Volume 58
Number 5
October 1991

# CONTENTS *continued*

**Paternalism and Drug Abuse** *Tziporah Kasachkoff* .................... 412

**Pharmacology of the Drugs of Abuse and the Development of Public Policy** *Robert B. Millman* .................................. 416

PUBLIC POLICY

**Drug Policy in New York City** *Stephen C. Joseph* ..................... 421

**Drug Use and Public Policy** *John Brian Murtaugh* .................... 423

**Drugs and the Social Body** *Patricia S. Mann* .......................... 427

**The Physician as a Servant of Public Policy** *Barry Stimmel* .......... 433

PANEL DISCUSSIONS

**Drug Use: Social and Scientific Background** *Moderator: Daniel A. Moros* ............................................................ 437

**Drug Use: Public Policy** *Moderator: Bernard Baumrin* ............... 441

ISSUES IN MEDICAL ETHICS

**Doctors' Dilemmas: Bioethical Challenges** *Moderator: Avi Barbasch Discussants: Richard Gorlin, Daniel Callahan, and Bernard Baumrin* ................................................ 443

**The Personal Ethics of the Physician: Curing Medicine from Within** *Edmund D. Pellegrino* ..................................... 452

**Medical Ethics: A Broader Definition** *Sidney M. Wolfe* ............... 455

## *Introduction*

Daniel A. Moros, M.D.
and
Rosamond Rhodes, Ph.D.

# Issues in Medical Ethics

As medical judgments and medical care increasingly influence our economic and social life, it has become necessary for individuals from different disciplines—physicians, lawyers, philosophers, legislators—to meet, discuss, and challenge each other's understanding of individual rights and professional and societal duties. Over the past two decades, consideration of such topics as the just distribution of social resources, patient autonomy, informed consent, physician obligation, and beneficence has been incorporated into the curriculum of most medical schools. Indeed, corresponding curricular changes have appeared in other professional schools as well as in graduate programs in the sciences. Some medical and scientific publications are also devoting more attention to moral issues.

This issue of the *Journal* offers discussions on ethics held at Mount Sinai under various auspices during the past three years. It contains adaptations of presentations at the 1989 and 1990 New York Regional Conferences in Medical Ethics sponsored by the Department of Medical Education of the Mount Sinai School of Medicine. (The proceedings of the 1987 and 1988 conferences were published in this journal in an earlier volume.) These conferences are designed to stimulate scholarly, but pragmatically oriented, debate on matters of immediate concern. Also in this issue are some thoughts of Edmund Pellegrino's on the physician's personal ethics; the 1991 Richman Lecture by Sidney Wolfe focusing on the social setting of medical practice and physicians' responsibilities; and a panel discussion, by Daniel Callahan, Bernard Baumrin, and Richard Gorlin chaired by Avi Barbasch, of ethical dilemmas physicians face in daily practice. All these papers address enduring issues and broad principles in the light of evolving problems reflecting the time of their delivery.

**The Morality of Drug Use.** Drug use in our society has become an issue of almost overwhelming concern, and its attendent problems have generated intense debate. The relationship of drug use to crime, disease, and child abuse fills our newspapers, magazines, and airways. Our presidents, governors, and mayors declare war on drugs, and indeed, the State Department, the Pentagon, and even the Central Intelligence Agency are involved. In the midst of this whirlwind of media blitz, community anxiety, and political clamoring, positions quickly harden in public, before being informed, and then rapidly become dogma. Coherence and consistency are sacrificed in this atmosphere of crisis, and public policies risk war with each other.

The clash between the need to protect our population from harm and our strong respect for individual autonomy is especially dramatic in the debate over drug control. The regulation of drug use within a society deeply committed to personal liberty demands that we thoughtfully confront the complex philosophic, legal, and scientific questions raised by these competing principles. Although drugs of choice and patterns of use do change, the questions that must underlie a rational drug policy remain. Among such questions addressed at the 1989 conference, "The Morality of Drug Use: License vs. Prohibition," were these:

- What justifies governmental intrusion into private drug use?
- What danger does any particular drug pose for the individual or for society?
- Should drug addicts be treated as criminals or patients?
- Can rational and consistent law be constructed to govern drug use?
- Should physicians' prescriptions for drugs be regulated or monitored?
- Should we continue to tolerate historically sanctioned and socially entrenched drug use?

Because these questions cut across many disciplines, this conference sought a productive interaction of practicing physician-researchers (Marc Galanter, Robert B. Millman, Barry Stimmel), a historian (Daniel M. Fox), a sociologist-epidemiologist (Don C. Des Jarlais), philosophers (Tziporah Kasachkoff, Patricia F. Mann), and makers and implementors of public policy (John

---

Daniel A. Moros is an Associate Professor of Neurology at The Mount Sinai School of Medicine (CUNY). Rosamond Rhodes is an Assistant Professor of Medical Education at The Mount Sinai School of Medicine and of Philosophy at Hunter College (CUNY).

B. Murtaugh, a New York State Assemblyman; Stephen C. Joseph, in 1989 New York City Commissioner of Health).

**The Ethics of Surrogate Decision Making.** Problems related to making decisions for others often confront the medical caregiver. Providing medical care for individuals of limited competence requires that we carefully consider the nature of surrogate decision making and the role of the surrogate decision maker. These issues are pressed on us not only because of our growing technological abilities but also by our increasing emphasis on patient autonomy and our growing concern with the rights of children and the duties of the state toward children. Among questions that warrant consideration are:

- To what extent can statements made under the pressure of current illness be accepted as reliable guides?
- How should reports of previous statements by patients who are now incompetent be used? Who should assess patient competence?
- What criteria are appropriate for assessing patient competence?
- Who should decide for whom?
- How should a surrogate make decisions for others?

The 1990 conference on surrogate decision making sought insights from philosophers (Dan Brock, Jonathan Moreno), jurists (New York State Court of Appeals Judge Judith Kaye), and representatives of patients' advocacy groups (Fenella Rouse, Society for the Right to Die; David Hansell, Gay Men's Health Crisis). We specifically tried to challenge their positions with a series of case presentations by practicing physicians from The Mount Sinai Hospital. The New York State proxy act, which became law in June 1990 and went into effect in January 1991, addresses some of Rouse's and Hansell's concerns. Nevertheless, their appeals and concerns serve as important arguments for the existence of such legislation and cautions against its abuse.

The multidisciplinary approach is the hallmark of the annual New York Regional Conferences in Medical Ethics. The objective of these conferences is to offer our immediate intellectual and professional community the opportunity for contemplative reflection on issues of urgent and enduring concern. Our aim is to question views which may have been conceived at too comfortable a distance from decision making or in the deceptive isolation of wrestling with only a single issue. We want to provide a forum in which those who will be obliged to carry out policy can participate in discussion with those whose primary responsibility is to formulate or critique it.

**Personal Ethics of the Doctor.** Edmund Pellegrino here offers six orienting points of reference to help doctors find their way to the moral path their lives should follow. He exposes the ethical pitfalls that lurk in the quandary-ridden practice of medicine today and implores new doctors to keep to the track required by their obligation to the high standard of professional responsibilities. His advice is a valuable tool for physicians at every stage in their careers.

**Doctors' Dilemmas.** Avi Barbasch leads an interdisciplinary panel—Richard Gorlin (medicine), Bernard Baumrin (philosophy), and Daniel Callahan (medical ethics)—in a discussion of some of the ethical issues doctors confront in their daily practice:

- Should a physician always do what is "best" for the patient?
- Does the physician have an obligation to act for the good of society?
- Is a doctor obliged to consider cost of treatment?
- When resources are scarce should age be considered in limiting treatment? Should character count?

**Social Responsibilities of the Doctor.** Sidney Wolfe, director of the Health Research Group, here confronts the uncomfortable issues of conflict of interests and finanical temptations and gain within the health care system—the medical profession, hospitals, pharmaceutical industry suppliers, and regulatory agencies. His central theme is that "physicians have a dual responsibility not only to practice and to research ethically, but also to spend . . . time on broader societal change."

We hope the discussions presented in this issue will stimulate additional thinking along these lines for those who did not have the opportunity to be present with us at these events.

# 1. Surrogate Decision Making

# Staking Out the Law

Hon. Judith S. Kaye

Any discussion about life and death, hopeless illness, right and wrong, and surrogate decision making is inevitably one of profound uncertainties. When we consider wrenching, often heart-rending subjects of this nature, there are many questions, but not many answers we can know with certainty.

So as the first discussant, with the privilege of defining my own topic, I begin the discussion of decision making and the law at a point of relative certainty, by summarizing existing law in this area—which in New York State is largely judge-made, rather than statutory, law. Stating what the law is seems to me a natural springboard for any discussion of ethics, which of course transcends the law.

I do not mean to suggest that we have the answers in the law either. Indeed, our common law system of justice—the law built by judges, one case at a time—has a vital component of uncertainty, which enables the law over time to respond and adapt to changing values and changing circumstances. To serve an evolving society, the law can never be rigid or immutable.

Yet a core of certainty is still necessary in the law. Cases decide disputes, to be sure, but court opinions—especially appellate court opinions—go beyond the parties to establish and define the limits of tolerable, or lawful, conduct for society at large. In this manner, hopefully, similar conduct will have similar legal consequences, irrespective of who you are, and individuals can fashion their business and personal affairs to come within the protection of the law. To a great extent, therefore, courts have both a tradition and a responsibility of stability. Courts simply cannot decide a question one way for one litigant, and then a different way in the next case.

I do not intend to present a lecture in jurisprudence. But because this discussion is interdisciplinary, I offer these prefatory thoughts both by way of background about my own discipline, and by way of explanation for my audacity in announcing that, even on deeply unsettling questions of life and death, the law offers a measure of certainty.

Further, limiting myself to stating what the law is gives me some comfort in participating in this lively and controversial discussion. Heightened consciousness of professional ethics being the order of the day for this seminar, I am sure readers appreciate that as a judge it would be inappropriate for me to offer my personal views on past decisions, or to express opinions that might betray a predisposition for future cases.

## What the Law Is

The bedrock principle in the law on medical treatment decisions is one of self-determination, or autonomy. Every person of adult years and sound mind has the final word on what should be done with her body and her life. Whatever the doctors may say, under the law it is the patient's ultimate right to agree or disagree, even as to life-sustaining treatment and care. Courts speak, often eloquently, of the concept of individual autonomy as a common-law right. It has sometimes been equated with a constitutional liberty interest—indeed, that view has just been put before the United States Supreme Court in the *Cruzan* case—and it has also been characterized as a constitutional right of privacy and due process. Autonomy is the basis of statutes that require informed consent for all sorts of medical procedures.

The right of every competent adult to refuse treatment is not absolute. It can be overridden by compelling state interests like preservation of the public health and prevention of suicide. But as a practical matter that right is very broad. Just this year, for example, our Court upheld the right of a Jehovah's Witness, after the birth of her son, to refuse blood transfusions believed necessary to

---

Adapted from the author's presentation at the New York Regional Conference on Medical Ethics at the Mount Sinai Medical Center on March 9, 1990. The author is a Judge of the New York Court of Appeals, the State's highest court. Address reprint requests to the author at Court of Appeals Hall, 20 Eagle Street, Albany, NY 12207.

save her life (1). The hospital urged us to adopt a balancing test, recognizing a state interest in preserving the life of a young, otherwise healthy mother facing a risk of death that could be eliminated by a simple blood transfusion—but we declined the hospital's invitation. Similarly, we have held that involuntarily committed mental patients cannot, simply by reason of their status, be presumed incompetent to refuse psychotropic drugs (2). Incompetency to manage personal affairs, in other words, is not necessarily incompetency to understand and appreciate information necessary to give informed consent for medical treatment.

Although medical treatment choices may be hard to make in fact, it is at least easy to state the governing principle in law. Autonomy controls. Paternalism, or beneficence—the ethical principle of helping others—has not been a relevant concern in the case of adults competent to make their own treatment decisions. We do not impose our notion of a good prognosis or worthwhile life on competent adults; that is a matter of their own subjective choice.

That same clear principle also controls medical treatment decisions by incompetent persons, although the mere statement of the proposition betrays its complexity. In particular, the discussion has been inordinately complicated in the past decade by the successes of medical technology in prolonging life, or as sometimes may be the case where there is incurable terminal illness, in just prolonging what has come to be known as "the process of dying." The growing elderly population and recent advances in life-sustaining technology have given rise to imponderable new questions for all of society.

When a person lacks capacity to decide, in New York State the case law establishes that medical treatment can be terminated or withheld only where there is clear and convincing evidence of the patient's own wishes (3). Although a writing is not required, we do insist on a high level of proof of the persistence and the seriousness of the patient's alleged subjective intention. The Court of Appeals has made clear that every effort must be exerted to minimize the opportunity for error, both in discerning the individual's true intention and in assuring that there has been no recent change of heart. Any error is to be made on the side of continuing life.

The Court has acknowledged its openness to two methods by which individuals, while competent, can indicate their preferences regarding future medical procedures. The first is the living will—a writing expressing an intention that at some stated point in the future the individual would or would not wish certain medical procedures. The Court has said that, in determining whether there is clear and convincing evidence of what an incompetent person would have wanted, it would give weight to such earlier written expressions. I should point out, however, that—unlike most other states—in New York we have no statute actually permitting living wills. Our legislature, year after year, continues to debate and defeat proposed legislation legalizing them.

The second method recently mentioned by the Court of Appeals is the health care proxy, or designation by a person while competent that another named person is to convey her medical treatment preferences in the event of incapacity. The Court took its cue from a recent New York statute allowing individuals to create powers of attorney that come into effect when the named person determines that the maker has become incompetent. In light of this statute, our Court said it no longer saw any reason why those wishing to appoint another to express their desires with respect to medical treatment, in the event they become incompetent, could not do so formally through a power of attorney.

Like living wills, health care proxies have not yet been sanctioned by the New York legislature; the subject is one of continuing debate, though the concept has been endorsed, after comprehensive study, by the New York State Task Force on Life and Law.

As a practical matter, living wills and health care proxies seem to be gaining in use, even without legislative sanction. I understand that many lawyers counseling clients about wills and estate planning matters now routinely ask about these additional documents. The tremendous publicity generated by recent court cases unquestionably has had the beneficial effect of forcing people of all ages to confront the prospect of their mortality in light of the wonders of modern technology, and to set down in writing what they want and don't want by way of future care. Although it seems that nothing on the subject of surrogate decision making is absolutely perfect—some practical or logical flaw can be found in every proposal—I have to think that a competent person's own serious written expression of intention for future care, made with the knowledge that it can at all times be changed, approaches the ideal.

With the addition of a "do not resuscitate" statute about a year and a half ago, in New York State this is about where the law stands on the subject of surrogate life-and-death decision making (4). In general, when a patient cannot herself exercise the right of self-determination, but her preferences can be clearly established from past acts and statements, a court will give effect to those preferences. In the only case to come before the Court of Appeals in which the patient's condition approximated a persistent vegetative state—the *Eichner* or Brother Fox case—we found sufficient indication that the patient had made known his intention so as to permit termination of life-sustaining apparatus (5). Last summer, in the *Elbaum* case, the intermediate appellate court in New York State, applying that same subjective judgment test, also reached that conclusion (6). By contrast, in the recent *O'Connor* case, our court concluded that the threshold level of required proof had not been met.

Our Court of Appeals has otherwise rejected the concept of surrogate life-and-death decision making because of a fundamental commitment to the notion that no person or court should substitute its judgment as to what would be an acceptable quality of life for another. As Chief Judge Sol Wachtler recently wrote in the *O'Connor* case: "Every person has a right to life, and no one should be denied essential medical care unless the evidence clearly and convincingly shows that the patient intended to decline the treatment under some particular circumstances."

In this vigorous young republic of 50 sovereignties, it should come as no surprise that the law differs among the states. Since the *Quinlan* case in 1976, several other states have embraced the principle of surrogate decision making for life-sustaining treatment, reasoning that incompetent persons should have the same freedom as others to decide whether they will undergo invasive medical treatment. Cases and statutes in other states have concerned themselves with defining the requisite incompetence, identifying the decision maker, and spelling out procedures for actually implementing surrogate decisions.

The Massachusetts court determined that an individual's right to privacy and autonomy could only be protected by the "detached but passionate" judgment of a court (7). It viewed family members as too emotional or too biased. The court said that it would make the decision that the incompetent would make if competent, "taking into account the present and future incompetency of the individual as one of the factors which would necessarily enter into the decision-making process of the competent person." Disclaiming any intention to make decisions based on life quality, the Massachusetts court set forth several "objective" guideposts it would use in decision making, such as the patient's expressed preferences regarding treatment, her religious convictions, impact on the family, and the prognosis with and without the treatment (8).

The New Jersey court—perhaps more than any other state court—has several times revisited the subject of surrogate decision making for incompetents, and it has evolved a continuum of tests, ranging from purely subjective to purely objective (9).

New Jersey's subjective test, like New York's, effectuates the patient's personal right of self-determination. Next on the continuum is a substitute-judgment test, which applies where there is inadequate evidence of the incompetent's wishes for a wholly subjective standard. The New Jersey court expressed a preference that close family members, rather than strangers or courts, act as substitute decision makers (or guardians) for the incompetent patient, both because of the special bonds they have with the patient and because they are most familiar with the patient's life—including philosophical, religious, and moral views; life goals; values about the purpose of life and the way it should be lived; and attitudes toward sickness, medical procedures, suffering, and death. The guardian is permitted to exercise the patient's right only after her hopeless illness had been confirmed by an ethics or medical-prognosis committee. So long as the prescribed procedures are followed, the court has emphasized that it is unnecessary to seek judicial authorization for particular decisions, and no liability will attach for a physician's conduct in discontinuing certain treatment.

Finally, the New Jersey court has a "best interests" test for patients who have never been competent or never expressed any preference while competent. Under this test, a guardian is empowered to authorize termination of medical treatment where for example the prognosis for the patient's continued existence is one of extreme pain, not outweighed by any other benefits of continued life. A guardian in such a case must seek a judicial determination of the patient's incompetency to make that decision, offering clear and convincing evidence of at least two physicians who have personally examined the patient. Petitions for the termination of life-sustaining treatment may be treated as potential cases of abuse, to be investigated by the Ombudsman for the Institutionalized Elderly. The Ombudsman gathers evidence from disinterested physicians, as well as from the patient's attending physician, all of whom must agree that the prognosis is hopeless. If all do concur, the guardian may order that life-sustaining treatment be removed, and there will be no risk of liability for that action.

In addition to these judicial models, there are by now eleven states that have, by statute, similarly authorized substitute decision making for incompetent, terminally ill adults as part of their "living will" or "natural death" acts (Arkansas, Connecticut, Florida, Iowa, Louisiana, New Mexico, North Carolina, Oregon, Texas, Virginia and Washington). The statutes generally provide that the patient's attending physician and at least one other specified person may make the decision to end treatment. As a precaution against abuse, most of the statutes require at least one physician in addition to the patient's own doctor to certify that the patient is suffering from a terminal condition or irreversible coma, or is being kept alive only by extraordinary means.

In broad outline, these are the various legal approaches to surrogate life-and-death decision making. Courts and legislatures that have adopted some form of substituted judgment have, additionally, gone to considerable lengths to reduce the uncertainties to specified criteria, with a view to minimizing abuse, protecting the decision makers from liability, and facilitating the process without the need for repeated court intervention. But instead of trying to summarize these further details, I'd like to return to where I began—which is the legal process generally, and particularly the role of the law in the issues under discussion.

### The Role of the Law

Speaking personally, I can say that, although every case is important, the life-and-death cases are an unparalleled part of a judicial docket—as well they should be. Courts traditionally have protected the rights of incompetents, and they deal every day with all manner of hard issues. But nothing comes close to the questions of living and dying. They are the most anguishing issues for judges—themselves parents, children, spouses, friends, mortal human beings.

Many times courts have voiced their discomfort with these cases. It's not because the issues are unpleasant or difficult; courts are at the center of some of the most troubling societal questions today, even sometimes criticized for deliberately thrusting themselves into that position. Moreover, the courts acknowledge their responsibility and function to resolve all controversies that come properly before them, and they approach these cases like the others—carefully, dispassionately, hopefully with wisdom and restraint.

The reservations courts have expressed are of entirely different origin. They grow out of a feeling that these cases are really not best suited to resolution by courtroom combat. Judicial proceedings tend to be long, formal, adversarial, and intrusive. They are financially and emotionally draining for the parties; family grief immediately becomes public property. By the time a case makes its way through the trial and appellate courts, nature may well have settled the immediate controversy. Once the patient dies, the litigation is technically moot, and the proceeding—if it continues—becomes a scholarly legal exercise rather than a flesh-and-blood case.

Then too, the acute importance of factual nuances in these cases creates a significant limitation, and I mean this in two ways. First, it may be difficult—perhaps more so than in other litigation—to replicate in the courtroom the critical reality that may become evident over time at the bedside. Second, a judicial opinion is inextricably tied to the facts of the particular case—subtle changes in the facts might affect the rationale, even the outcome—restricting applicability to other cases with different facts. These are all innate qualities of the judicial process.

Public policy pronouncements on issues of wide social significance are often better made by legislatures, whose function it is to promulgate broadly applicable rules. Unlike courts, legislatures can initiate public hearings and appoint independent interdisciplinary commissions or task forces, which can study a subject in all of its permutations and make recommendations as to the best course of action for an entire range of interests. And what could be more fundamentally a matter of public policy than the value this society chooses to place on continuing life?

From the relative paucity of "right to die" cases in the courts—there are no more than a few dozen such reported cases throughout the nation—it is also obvious that these decisions have traditionally been made elsewhere, in hospitals and other health care facilities, by physicians and families, quietly and cooperatively, outside the legal arena. I have no doubt that this continues to be true. Only the rare case ever reaches a court, and even these singular, widely publicized events tend to be petitions for guidance rather than damages actions or criminal prosecutions for what a doctor has done; it is obviously not fear of personal liability that drives a case of this nature into court.

So I ask: why am *I* here, as part of this program? How have we all arrived at this point, as limited partners in life-and-death decision making? Has something gone wrong—or right?

Courts and judges have really become a part of intensive-care decisions only within the past decade. For the first time in our nation's history, a "right to die" case has reached the United States Supreme Court. It is clear that this was inevitable. There are some good reasons and some bad reasons why even these intensely private, intensely individual matters are lumbering their way through the halls of justice, fortunately stirring up much-needed public awareness every step of the way. Legislative inaction and general societal litigiousness are surely among the reasons. More than a century ago, Alexis de Tocqueville observed of the American political system that the great issues are all ultimately resolved in the courts.

But the principal motivating factor, I believe, for resort to the courts in these life-and-death cases has been the technology that has dramatically changed the health care options, scientific advances making death more and more often the conscious act of a person who can prolong the fact of life but not its dignity or quality. New uncertainties of a vastly different dimension have been added. It is hardly surprising, therefore, that the courts have been brought into the process, as objective, impartial, responsible arbiters lending guidance and support to the day-to-day decision makers, by drawing lines and establishing certain outside boundaries. I have little doubt that this will continue.

From time to time in the course of my own judicial duties, I consult the writings of Benjamin Cardozo who, among other great achievements in the law, served as Chief Judge of the New York Court of Appeals. His wisdom and prescience are astounding. Judge Cardozo explained nearly 70 years ago—long before the sophisticated technology that brings us together today—that judges have a social mandate to distill the values and mores of the day. In so doing, he wrote, their conclusions necessarily must "be subject to constant testing and retesting, revision and readjustment; but if they act with conscience and intelligence, they ought to attain in their conclusions a fair average of truth and wisdom" (10).

In the end, I have to admit that I am in this company for wholly illegitimate reasons—not at all to educate *you* about the ethics of surrogate decision making. I undertook to summarize the existing law, including its role and innate limitations. But mostly I came to listen. It is clear to me that, for the future, if judges in their decisions are indeed to attain "a fair average of truth and wisdom" on this extraordinary subject, we can only gain from listening to all of you.

### Postscript

That public consciousness has been raised to issues of medical ethics, particularly "right to die" issues, is amply demonstrated by the many programs and writings that have appeared on the subject since March 9, 1990. Several have centered on two long-awaited legal developments: the *Cruzan* decision issued by the United States Supreme Court on June 25, 1990, and the new New York State health care proxy law, effective January 18, 1991. Although both developments deserve extended consideration, at least a brief reference is necessary to round out this survey of the law.

For most everyone, the five-four *Cruzan* decision was both encouraging and disappointing—whatever one's point of view. For the Cruzan family, of course, the decision was nothing but a crushing disappointment, for it denied them the right to discontinue the feeding and hydration tubes that have sustained Nancy Cruzan for more than seven years in a persistent vegetative state. The court upheld the right of the State of Missouri to continue artificial nutrition and hydration because there was no clear and convincing evidence that Ms. Cruzan herself would have wished otherwise.

Those advocating that the Supreme Court "get into the business" of resolving these issues—to settle them on a nationwide basis as a matter of constitutional due process of law—took heart from the court's suggestion, and assumption for this case, that the forced administration of life-sustaining medical treatment might be a constitutional "liberty" interest. By the same token, those opposing Supreme Court constitutional involvement celebrated the fact that the case was actually decided on the narrowest possible ground, with many anguishing issues explicitly reserved for another day. For the moment, therefore, the only certainty yielded by the Supreme Court's first foray into right-to-die issues is that state courts and legislatures will continue much as they have in the front lines of shaping the law in this area.

Six days after the *Cruzan* decision, the New York State legislature passed the health care proxy law (N.Y. Public Health Law, Article 29-C, §§ 2989–2994).

The new law, in large measure the 1987 recommendation of the Governor's Task Force on Life and the Law, for the first time allows adults to designate, by a properly drawn proxy, their agents to make medical treatment decisions in the event they can no longer do so themselves. Unless otherwise specified, the agent will have the same authority to decide treatment as the principal when competent. Good faith reliance on the specified agent's choices will in general protect health care professionals and facilities against liability. Significantly also, the new law specifies that it is to be provided in summary form to health care facilities' staff, as well as to patients and family members at admission—which should prompt forthright discussion at that critical time.

Plainly this new statute, the recent decisions, and all the media attention will help to stimulate public awareness and thereby advance the concept of self-determination in medical treatment decisions.

## References

1. Matter of Fosmire v. Nicoleau, 75 N.Y.2d 218, 551 N.E.2d 77, 551 N.Y.S.2d 876 (1990).
2. Rivers v. Katz, 67 N.Y.2d 485, 495 N.E.2d 337, 504 N.Y.S.2d 74 (1986).
3. See Matter of Westchester County Med. Center (O'Connor), 72 N.Y.2d 517, 531 N.E.2d 607, 534 N.Y.S.2d 886 (1988).
4. N.Y. Pub. Health §§ 2960–78 (McKinney Supp. 1990). In 1985, the Legislature added article 80 to the Mental Hygiene Law (made permanent this year), to provide an alternative health care decision making procedure for residents of mental hygiene facilities operated or licensed by the State who cannot make such decisions for themselves, and for whom there are no available or willing decisionmakers. Volunteer surrogate decision-making committees can decide whether the patient has the capacity to consent and, if not, whether proposed treatment is in the patient's best interests, taking into account the known beliefs, preferences and values held by the individual. (N.Y. Mental Hyg. Law §§ 80.01–.13 [McKinney Supp. 1990].)
5. Matter of Eichner (Fox), 52 N.Y.2d 363, 420 N.E.2d 64, 438 N.Y.S.2d 266, cert. denied, 454 U.S. 858 (1981).
6. Elbaum v. Grace Plaza of Great Neck, 148 A.D.2d 244, 544 N.Y.S.2d 840 (1989).
7. Superintendent of Belchertown State School v. Saikewicz, 373 Mass. 728, 759, 370 N.E.2d 417, 435 (1977).
8. See Rogers v. Commissioner of the Department of Mental Health, 390 Mass. 489, 506, 458 N.E.2d 308, 318–19 (1983).
9. See e.g., In re Jobes, 108 N.J. 394, 529 A.2d 434 (1987); In re Peter, 108 N.J. 365, 529 A.2d 419 (1987); In re Conroy, 98 N.J. 321, 486 A.2d 1209 (1985); In re Quinlan, 70 N.J. 10, 355 A.2d 647, cert. denied 429 U.S. 922 (1976).
10. Cardozo B. The nature of the judicial process. New Haven: Yale University Press, 1921, 135–136.

# AIDS and Surrogate Decision Making

David A. Hansell, J.D.

In this article, I present my perspectives on surrogate decision making from the vantage point of one who works with persons with AIDS and AIDS-related illnesses.

I provide some context for my observations by explaining a bit about the legal services program at Gay Men's Health Crisis (GMHC), then telling three stories that reflect experiences our clients have had and illustrate some of the critical problems in the current legal structure governing advance health care planning.

The legal program at GMHC, which is part of the umbrella of psychosocial, educational, and advocacy services offered to clients, is dedicated to helping persons who have either AIDS or HIV illness to maintain full control of their lives and full exercise of their rights. Like any legal services program, we provide individualized services and handle discrete legal matters. In our case, these range from drafting wills, powers of attorney, and medical directives, to representing clients in areas as diverse as housing, insurance, discrimination, family law, immigration, and debtors' rights. But we are also committed to a concept I would awkwardly describe, for want of a better phrase, as "preventive law." We try to help our clients identify, anticipate, and avoid legal complications through advance planning in many areas. None of us, regardless of our health, needs the stress or unpleasantness of battling the legal system, but people with life-threatening illness need it least of all. Advance health care planning, of course, plays a central role in this mission.

Yet as much as we encourage people to plan earlier for medical decision making in the event of future incapacity, our advice must be hedged with numerous caveats and qualifiers, reflecting the limitations of the law in this area, limitations that in some ways are particularly troubling in the AIDS context. Here are some examples.

Case 1. The first involved a client I will call Peter, who was hospitalized with severe and untreatable human immunodeficiency virus dementia. He had not executed documents of any kind prior to becoming mentally incapacitated, and by the time we were called by his lover of many years, John, Peter was clearly incapable of understanding or executing a power of attorney, living will, or medical directive. His attending physician felt that there was no medical reason for him to remain hospitalized, and that any treatment he needed could be provided in another setting. John's wish was to bring Peter home to their shared apartment, where he could provide all necessary care with the assistance of a daytime home health aide to cover the times John was at work. Peter had no medical insurance to cover this, but had financial resources that exceeded Medicaid levels and were sufficient to cover home care costs for some time. The hospital's social worker felt that this was the most appropriate discharge plan.

Peter's family, however, from whom he had been estranged but who appeared on the scene while he was hospitalized, objected. They urged that Peter be institutionalized in one of the chronic-care institutions run by the City of New York, such as Coler or Goldwater Hospital. It was not clear whether their objections stemmed from their feelings about Peter and John's relationship, or from the belief that Peter's assets would be conserved if expensive home care were avoided. Whatever the case, the hospital refused to discharge Peter to his home in the face of his family's objections, despite the social work and

---

Adapted from the author's presentation at the New York Regional Conference on Medical Ethics at the Mount Sinai Medical Center on March 9, 1990. From the Legal Services Department, Gay Men's Health Crisis, New York, NY. Address reprint requests to the author, Deputy Executive Director for Policy, Gay Men's Health Crisis, 129 West 20th Street, New York, NY 10011.

medical staff's clear statements that a home setting was preferable and that John was a willing and able primary-care provider. A conservatorship battle ensued, but Peter died in the hospital before it was resolved.

Case 2. A second, related incident involved a client I will call Robert, who called me in tears one morning from a hospital pay phone. He and his lover, Joseph, were both diagnosed as having AIDS, and his lover was in extremely serious condition. Joseph had executed a medical power of attorney naming Robert as the person he wished to make health care decisions for him in the event of incapacity. Nevertheless, at the time Robert called me, Joseph's family was closeted in his hospital room with his physician making a decision about whether to disconnect the respirator. The family had barred Robert from the room, and the hospital staff had been unwilling to intervene even when presented with the medical power of attorney. Although I offered to get in touch with hospital's legal staff immediately, Robert felt he was not up to such emotional confrontation and, although he felt he was betraying his lover's wishes, decided simply to leave matters alone. Joseph died shortly thereafter.

Case 3. Another situation reached judicial resolution in a case called *Evans v. Bellevue Hospital*. Our client, Tom Wirth, had executed a living will and a medical power of attorney naming his friend of 22 years, John Evans, to make all medical care decisions. Tom was hospitalized with probable toxoplasmosis, and the hospital physician wanted to begin antibiotic treatment to control it. John maintained that Tom had given clear indication that he did not consider his overall medical state to be a "meaningful quality of life," and hence that he intended in his living will to reject life-sustaining treatment for this opportunistic infection because there was no reasonable possibility of recovery from his AIDS-related condition. The hospital, however, insisted that there was a reasonable likelihood that Tom would recover from toxoplasmosis, and hence that the living will did not reject that specific treatment. The State Supreme Court agreed and refused to enjoin further medical treatment.

**The Medical and Social Realities.** What do these apparent failures of surrogate decision making tell us about the problems we must address? Before we answer that, I think we must situate our discussion within an understanding of the medical and sociologic realities of the AIDS epidemic.

First, issues of surrogate decision making may differ depending on whether they arise from a sudden accident or injury or from a chronic disease. My clients are rarely without reason to anticipate the possibility of becoming incapable of medical decision making, and hence without some cause to reflect on their feelings about life-sustaining treatment in extremis. This does not mean that all (or even most) people with AIDS actually do what can be done to anticipate and control the course of future events. These decisions, which ultimately involve an unblinking confrontation with one's own mortality, are probably the most difficult that human beings are forced to tackle. But one of the things that never ceases to amaze me in doing this work is the profound courage and strength of character with which most of the people I see do confront and resolve them.

Moreover, the nature of AIDS as an illness enhances the complexity of advance decision making. AIDS is a disease with many and unpredictable manifestations, including many involving neurologic complications that may affect decision-making capacity at a stage when death is not imminent. Furthermore, many of the treatments for AIDS-related infections are at experimental stages, and some that are not involve serious toxic side effects that may militate against their use even if they would be narrowly effective against a particular manifestation of the disease.

Thus, in the AIDS context at least, the discussion of surrogate decision making is certainly not an ethicist-inspired euphemism for the issue more crassly articulated as whether to "pull the plug." A host of complex and weighty medical decisions must be made in treating a patient with AIDS, and often no standard course of treatment with demonstrated efficacy exists. Not to decide thus becomes a decision in favor of nonaggressive (as opposed to aggressive, albeit often experimental) treatment.

Sometimes an opportunity arises to explore with the individual in some detail and in advance the likely courses that the disease may follow and the types of treatment decisions that may have to be made. Yet even under the most ideal circumstances, it is not possible to anticipate and consider every treatment decision that may arise in the course of an illness. Nor can every decision be so neatly characterized as one of whether to provide or withhold life-sustaining treatment, irrefutably resolvable in favor of life sustenance absent the patient's clear expressions of intent to the contrary. The reality, as we all know it, is that substituted judgment is, and must be, exercised all the time. The real question is not whether the judgment of another can be substituted for that of a patient lacking capacity, but rather *whose* judgment is substituted. Should it be the attending physician's, the hospital counsel's, or the next-of-kin's as our law inflexibly defines it? Or should it be the patient's right to make this determination? Whatever the answer—and mine is quite clear—we cannot delude ourselves into believing we can avoid the question.

This is where the sociology of the AIDS epidemic comes into play. We all know that the disease affects many in nontraditional living situations, many whose definitions of family may not fit within conventional legal categories. As the New York State Court of Appeals recently acknowledged in the area of tenancy succession rights, I believe we must recognize that it is the reality of family relationships, not externally imposed definitions, that should govern. Yet the practical reality is that when critical medical decisions must be made, hospitals generally look to statutorily defined next-of-kin to make them. Whether this reflects the internalization of social biases that fail to recognize the diversity of human experience or a more calculating concern about financial liability, the result is that the patient's control over his or her own destiny is undermined.

Control is of vital concern to people with AIDS, largely because the disease inherently involves such a profound loss of control—over one's physical health, one's social world, one's emotions, one's finances, and so on across the board. But another critical factor stems from the fact that most of those affected by AIDS are gay men and people of color—groups long marginalized in our society. There is no question that AIDS, particularly the inadequate and moralistically driven social and governmental response to AIDS, has enhanced the sense of expendability among the communities bearing the brunt of the epidemic. It is difficult to convey adequately the sense of fear and anger that is generated when, for example, a President of the United States refuses for six years to acknowledge even the existence of an epidemic that is killing thousands of American citizens—not, however, citizens who play any significant role in his political or moral calculus.

What this means is that many of those affected by AIDS have little reason to trust institutions, medical or otherwise, to safeguard their interests or even to care about their survival. There are, of course, countless medical practitioners of enormous caring and sensitivity who are striving in the most unselfish way to respond to this epidemic. But it may be little comfort to many gay people to know that life-and-death medical decisions are in the hands of such institutions as, say, the Catholic Church, or even in the hands of family members from whom they may be estranged because of their sexuality. Perhaps even more than others, people with AIDS have reason to be vitally concerned that these decisions be entrusted to others of their choice.

What are the implications of all of this for the law of surrogate decision making? Most fundamentally, we must abandon the fiction that we can do without a mechanism for the exercise of substituted judgment, and we must have a vehicle that allows for the designation of another person to make medical decisions in the event of incapacity. It is simply immoral to leave these decisions to be quietly made in hushed discussions between a physician and the person seen as most likely to sue if not consulted. It deprives many of us of perhaps the most fundamental element of our personal autonomy—the ability to decide in whose hands these decisions should lie.

We also need a vehicle for expressing, simply and specifically, intentions about future health care. As *Evans v. Bellevue Hospital,* together with the *O'Connor* decision, made abundantly clear, specificity in the expression of one's wishes about life-sustaining care is essential, and I believe appropriate. I am less concerned about whether the "clear and convincing evidence" standard represents an excessively high hurdle than about whether people will be dissuaded from making any written statements of intent at all by any confusion in the law and by the lack of easy mechanisms for doing so. We should all, I think, want surrogate decision making to be guided decision making to the greatest possible extent, and we must use the law to facilitate that.

Thus, in concrete terms, I believe that health care agent legislation is the proper course to follow in responding to the problems I have tried to elucidate. I am also pleased that the New York law establishing health care proxy procedures does lay out a form for a health care proxy, which allows for the appointment of an agent and for specific statements about wishes or intentions. I suspect we may ultimately need to move toward even greater specificity in that regard, but that can evolve with time.

A final concern has to do with the obligation of a hospital or other health care facility to respect the wishes of a patient. Any vehicle designed to accomplish the goal of patient control over future decision making will fail if it is not enforceable. I am quite concerned that legislation not make it all too easy for a facility to refuse to carry out the wishes expressed in a health care proxy, by the simple expedient of throwing the matter into the courts. The kinds of decisions with which we are concerned are generally decisions that must be made quickly, and delay in carrying them out would vitiate the basic purpose of empowering a surrogate: to make the decision that the principal would make if he or she were capable. A health care facility should not have carte blanche to avoid carrying out the agent's instructions simply by bringing a proceeding to challenge his or her authority, or to challenge the validity of the proxy. An essential is a threshold requirement that the facility has cause to believe the proxy is invalid or is being utilized in bad faith; another is sanction against the facility for abusing the special proceeding as way of undermining the agent's authority.

These are my most general concerns. I do believe that health care agent legislation is a significant gesture of respect for the dignity·and autonomy of those who are fighting the lonely battle against AIDS or other terminal diseases. The battle is ultimately a lonely one, and the least that we as a society can do is to respect the wishes and beliefs of those whose lives are on the line, rather than to impose external and alien notions of moral order upon them.

# Surrogate Decisions in New York State on Sustaining Life

Fenella Rouse, J.D.

In discussing surrogate decision making, I have chosen to look at the doctrine from the point of view of the people involved in the situation that made the case law currently guiding us in New York. Two million people die every year in the United States, and many of those deaths involve decisions by surrogates to give, or not to give, some form of medical treatment. The question I raise is whether the current New York law is practical: does it in fact offer guidance that can be used in the majority of situations? If, as I believe, it fails to do that, we must ask whether it is either legally or morally relevant.

**The O'Connor Case.** The present law on decisions to withhold or withdraw life-sustaining treatment is found primarily in the Court of Appeals' decision in *In re Westchester County,* and the legislation on orders not to resuscitate (DNR).

The patient whose care was at issue in *In re Westchester County* was Mary O'Connor, a woman of 77 with multi-infarct dementia as a result of a series of strokes. In the summer of 1988 she was a resident in a nursing home, where her daughters had placed her after she became too sick for them to care for her adequately at home. Both daughters are licensed practical nurses and had cared for their mother at home for several years. It was when they could no longer move her that they arranged for her care in the nursing home where one of them worked.

She was admitted to Westchester County Medical Center in June 1988, for treatment for sepsis, pneumonia, and dehydration, and was treated successfully with antibiotics and intravenous fluids. She was minimally conscious but unable to communicate competently. Her physicians recommended nasogastric tube feeding because she could not swallow.

When told of the recommendation, Mrs. O'Connor's daughters made the decision they were convinced she would have made herself and, on her behalf, refused consent to insertion of the nasogastric feeding tube. They made this decision based on everything they knew of her, including statements she had made over the years about medical treatment.

Mary O'Connor was not a novice in the field of health care. She had worked for many years as an emergency room clerk, and had cared for her husband and two brothers during their final illnesses. Her daughters reported a number of conversations in which she had said that she would not want to go on living if she could not take care of herself and make her own decisions. She had also said that she would never want any sort of intervention, any sort of life-support systems to maintain or prolong her life. She had never specifically mentioned feeding tubes, or any other particular treatments.

Based on these remarks and everything else they knew about her, her concededly devoted family decided that their mother's current condition was the one of which she would wish to say "no more." Her daughters, putting themselves in her shoes, were convinced that she would refuse tube feeding if she could. In other words, they made a surrogate decision for her.

Mrs. O'Connor's case went before the hospital ethics committee, which reportedly decided that it would be unethical to withhold tube feeding. A second, and different, surrogate decision was

---

Adapted from the author's presentation at the New York Regional Conference on Medical Ethics at the Mount Sinai Medical Center on March 9, 1990. From the Society for the Right to Die, New York. Address reprint requests to the author, Executive Director, Society for the Right to Die, 250 West 57th Street, New York, NY 10107.

therefore made by the committee, and the hospital informed the family that it would seek court authorization for providing the treatment.

The trial court heard all the evidence, including descriptions of Mrs. O'Connor's medical condition and her daughters' reports of her statements. The trial court then ruled that nasogastric tube feeding should not be instituted because there was, as required by New York law, "clear and convincing evidence" that Mrs. O'Connor would not want life-prolonging medical treatment in her current condition. The trial court's decision not to order treatment could be described as another surrogate decision on Mrs. O'Connor's behalf, although lawyers would prefer to say that the court was merely refusing to order treatment because, as a matter of law, the evidence clearly demonstrated that Mrs. O'Connor had, when competent, decided that she would not want to receive this treatment in these circumstances.

The hospital appealed the decision, but the Appellate Division agreed with the trial court. Again the hospital appealed, and some two months later the Court of Appeals, the justices of which had, of necessity, never met the people involved, reversed the trial court's decision and ordered that tube feeding be provided. Judge Kaye describes the court's ruling. In essence, it held that the evidence did not demonstrate that Mrs. O'Connor had herself made an informed decision in advance to reject tube feeding, and that no surrogate could make that decision for her. As a result, Mrs. O'Connor was given a feeding tube and was released from the hospital to a nursing home.

Other surrogate decisions were, however, made on her behalf. When Mrs. O'Connor was first in the hospital she had a DNR order, which was rescinded when her case went to court. She also had one when she was later in the nursing home. Mrs. O'Connor assuredly had not talked any more specifically about resuscitation orders than she had about feeding tubes. A statute in New York provides that DNR orders can be authorized by family and close friends when the patient is permanently unconscious or terminally ill, or when resuscitation would be futile. These are clearly surrogate decisions, laden with interpretations of what constitutes futility and a terminal condition. I say this advisedly, because although the statute that authorizes such decisions defines all these terms, law is imprecise in its own way, particularly when applied to changing medical conditions. The point is that some surrogate decisions have been approved by statute but not others, and I would suggest that what we are doing in New York is making a decision on behalf of others that in some circumstances, continued treatment may not be in their best interests.

Note also that the medical order written for Mrs. O'Connor at the nursing home after the Court of Appeals had ruled was for "supportive care only." In other words, despite the Court of Appeals ruling that there was no clear evidence about what she would want in this condition, it was agreed by the nursing home staff that some forms of life-sustaining treatment would be inappropriate. This was, of course, a surrogate decision, but because it was apparently consistent with the beliefs of those providing her care, it was made without legal challenge. Mrs. O'Connor died about a year later, with the feeding tube still in place.

I think we should ask why we are comfortable with making some decisions to withhold life support but not others. We should also ask why Mrs. O'Connor, and others like her, are held to an extraordinarily exacting legal standard that bears little relation to everyday life. One could argue that we are in fact abdicating our moral responsibility by saying that everyone in New York is required to receive all available life-sustaining treatment unless those involved with them can prove legally that they previously decided not to have it. Insisting that a patient receive treatment, against all the information we have from those who care about the individual, may in fact make her the symbol of our cause—our cause being our own sense that we are a moral society. We sacrifice the individual Mrs. O'Connors to a thoroughly well-intentioned but abstract desire to show that we care.

***Cruzan v. Harmon.*** In *Cruzan v. Harmon,* the patient was Nancy Cruzan, who was 32 and had been permanently unconscious for seven years, without any cognitive reactions and with no hope of recovery. [The U.S. Supreme Court decided the case in June 1990.—ed]

An estimated 10,000 people in the United States are in the same condition. For one, Karen Ann Quinlan, whose plight caught the country's attention and whose case made law 15 years ago, was in the same condition. Ten thousand is a large number, but it is a tiny proportion of those who die each year. And yet it is permanently unconscious people whose cases have made much of the law, rather than patients in Mrs. O'Connor's condition, which is probably one that much more commonly precedes death. The Missouri Supreme Court's decision in *Cruzan,* which followed that of the New York Court of Appeals by about a month, reflected the New York court's reasoning. Family members and friends testified in the Missouri trial court that she had spoken, both generally and in particular, to indicate that she would not want to be maintained by medical treatment if she were permanently without thought or feeling. The trial court found the evidence clear and convincing. The Missouri Supreme Court, however, characterized it as "woefully inadequate," and even suggested that not only was there no informed refusal in advance by Nancy, but that, in fact, no one could ever make such a decision in advance about future hypothetical situations. The Missouri court acknowledged the tragedy that had befallen Nancy and her family, but stated that as a court of law, it had no alternative but to apply the law, and that law provided no answer for the situation but to continue treatment.

Joe Cruzan, Cruzan's father, who had gone to court because he had been told that it was the only possible avenue of relief, expressed his feelings on the day that the United States Supreme Court heard his daughter's case. He said that for years, before he had finally gone to court, he felt as if he had been trying to start a car which would not budge. Now, he said, he felt as if he were running along behind the car with no idea where it was going, unable to stop it. After sitting in the Court and hearing the lawyers' arguments, he said he felt that the dispute was no longer anything to do with Nancy, him, or his family. The debate was being conducted in a language he could not understand, with technical rules that had no meaning for him. I would imagine that his feelings are shared by many of those who either petition the courts or attempt to apply the rules of law we have in some states to the actual cases of seriously ill and dying people.

**Surrogate Decisions in New York State.** I believe surrogate decisions about life-sustaining treatment are being made in New York, one way or another, with a good deal of frequency. Most people do not write their wishes out in advance. Only about 15% of adult Americans have living wills or durable powers of attorney for health care. Many families in New York must face the situation faced by Mary O'Connor's daughters. I know from my own experience that many like her are receiving treatment their families say they would not want. Some surrogate decisions not to resuscitate are made under the authority of the legislation on orders not to resuscitate; some surrogate decisions not to provide other kinds of treatment are made because families are able to report clear decisions to that effect made in advance. Other surrogate decisions not to treat are made because families and health care providers all believe it would be best not to treat. Surely these latter are ethically correct, but they may not be consonant with the law.

Surrogate decisions to provide treatment are often made in the same way, that is, because all involved believe it is best. Others are made when only the health care providers believe it is best, and the families either are not informed of their options, or do not have the necessary evidence or the financial and emotional resources to carry them through what might be a prolonged court battle. A few surrogate decisions to continue treatment are probably made when all involved think it would be best to stop, but are aware that New York law may require them to treat because of lack of evidence of the patient's wishes. These cases, however, are probably the minority. In my experience treatment will often be ended if the health care providers agree, regardless of the law. Similarly, it will often be continued if the health care providers think it should, regardless of the patient's wishes, unless they are in writing.

Is there a solution? There is, and it is one that other states have already adopted. Health care proxy legislation in New York [in effect January 1991—ed] would allow us to choose a surrogate to make all medical decisions for us, when we cannot act for ourselves.

Since many people will still not appoint proxies, we also need additional legislation [the Uniform Rights of the Terminally Ill Act?—ed] which lists those family members and others authorized to make decisions for people who have never indicated what they would want. Legislation of this sort exists in a number of states, and every indication is that it is being used wisely and well. It is providing a rule of law that makes practical, everyday sense and allows patients, their families, and their medical teams to retain some privacy as difficult decisions are made.

# The Physician's Experience:

## Infants and Children

Ian R. Holzman, M.D.

**Case Report.** A male baby was born 12 weeks early and weighed 1100 gm (2¾ lb). He suffered severe birth asphyxia and had Apgar scores of 2 (1 min) and 3 (5 min) in spite of immediate resuscitation in the delivery room. He was taken to the neonatal intensive care unit and placed on a respirator. Examination revealed an extremely neurologically depressed infant. He had multiple features consistent with a diagnosis of the Vachterl Association. These included deformed and foreshortened arms bilaterally (radial aplasia), anal atresia (lack of visible anal opening), esophageal atresia with a tracheal fistula (esophagus not connected to stomach, connection of distal esophagus to trachea), bilateral enlarged kidneys, and a heart murmur.

His chest radiograph revealed an enlarged heart and extremely severe hyaline membrane disease as well as some deformities of his spine. A sonar of his kidneys revealed moderate hydronephrosis bilaterally (enlargement of urine collection system, suggesting obstruction). A blood test indicated the lungs were not successful at supplying oxygen to the baby or removing carbon dioxide. Results of immediate bone-marrow chromosome analysis were normal.

The father of this infant immediately requested that no further care be given to the infant because of the birth defects. The mother did not initially express any opinion.

**The Mother's View.** The mother of the baby boy described her experience as follows (*reprinted with permission from the Pittsburgh Post Gazette, Nov. 3, 1984*):

> I opened my eyes and looked up into the face of a stranger with kind, sad eyes. He was speaking to me and as I struggled out of the anesthetic I heard the words, "No dramatic measures will be taken to save your child." Suddenly I realized where I was, and the events that had brought me to this moment came thundering into my mind.
>
> I had married my husband five and a half years before. The last four years had been spent with fertility specialists due to some unexplained problem with achieving pregnancy. None of the doctors could find a medical problem. This was before scientists had determined that extended use of the pill could delay pregnancy up to three years.
>
> The emotional component of infertility had sent me to seek support from a wonderful national self-help group for infertile couples. The trauma had been so dramatic for me that I had reached a stage of depression where I stopped going to work or even getting out of bed. At that point my husband located a counselor who specialized in infertility.
>
> After a number of extremely productive sessions, she encouraged me to come to a group session. This I resisted, as I am not a "joiner," but eventually my counselor convinced me that it was the best thing for me emotionally and financially.
>
> I went with trepidation and discovered a caring, intelligent group of women with enormous insight into the confusing emotions of infertility; anger against your own body, against fate, against God, against women who are pregnant by choice and fury at women who are pregnant by accident, the "why me?" syndrome, the added complication of adoption becoming more difficult, and all the rest. This initial loosely knit group grew into the first chapter which serves couples, not just women.
>
> Finally, I had achieved some control over the seemingly "uncontrollable" events in my life. I had decided to explore all options in infertility until the age of 35 and then to have a tubal ligation and become a serious career woman. (At that time, I had a "paycheck" job . . . no satisfaction but a healthy paycheck every two weeks.) Each month was still two weeks of hope and two weeks of depression, but the cycles weren't so violent.

---

Adapted from the author's presentation at the New York Regional Conference on Medical Ethics at the Mount Sinai Medical Center on March 9, 1990. From the Department of Pediatrics, Mount Sinai Medical Center, New York. Address reprint requests to the author, Department of Pediatrics, Box 1198, Mount Sinai Medical Center, One Gustave L. Levy Place, New York, NY 10029.

Then, in January of 1979, after several operations, a brief try at a mild fertility drug, clomid, and temperature-taking ad nauseum, during a lull in treatment came the news, "Yes, . . . your pregnancy test was positive!"

Our friends reacted to the news as second in importance only to the Second Coming. We were the most celebrated, pampered pregnant couple in town. (I secretly suspected it was largely because they were sick of hearing me complain.)

The pregnancy was uncomplicated initially, except that while only gaining 10 pounds I was enormous. The doctors suggested I had confused the dates, but I quickly reminded them that a woman under infertility treatment for four years would never "mistake" her dates. At seven months they did a sonogram and an amniocentesis to verify the weeks of pregnancy and whether the lungs were ready. My dates were right; the lungs were immature.

The hospital I had chosen did not have a neonatal intensive-care unit. My doctor, a young compassionate man, insisted that I go to [another hospital] when the amniocentesis caused premature labor.

We were whisked off in an ambulance and soon were among strangers. My water had broken and suddenly I was no longer enormous. I had little more than a pot belly. My husband and I were terrified. Suddenly, with shattering intensity, the possibility of something going wrong had intruded on our happiness.

The hospital was a wonder of modern technology. They stopped labor and gave me medication to mature the baby's lungs. The following night, a Tuesday, I went into labor again. After 5½ hours of labor, during which they did everything to me, including standing me nearly on my head to help the baby's weakened heart, the sounds from the heart monitor stopped.

With speed and drama I had previously seen only on TV, I was moved from the delivery table to an operating table. In that few seconds I realized that what I had wanted to be controlled by myself and the medical community was truly in the hands of God. I had time only to say the first few lines of the 23rd Psalm and I was out. I woke briefly in the operating room and asked if the baby was alive. They said he was, and I drifted off.

The next time I awoke I was facing the kind man with the deep, reassuring voice. He explained that he was head of a department dealing with complicated deliveries. He explained about oxygen deprivation, multiple birth anomalies including major heart defects, and concluded by saying to correct the physical defects alone would take two years of hospitalization, at the end of which the baby would still be severely retarded and probably have cerebral palsy as well.

Then he said that, because of this, no dramatic measures would be taken to save my child's life, and no measures would be taken to end his life.

At that moment my earlier realization returned. Faith, which until then had been an intellectual exercise, became reality. I felt in my heart that God was not a concept but a living presence. The cold fear left and I felt peace. The peace that comes from knowing, not in my mind but in my heart, that God is present and has a plan. I closed my eyes and delivered the fate of my son, David, into God's hands. He chose to take David after he had been with us for only eight hours. The pediatrician said his heart failed and attempts at resuscitation had not been successful.

The following morning was my 32nd birthday. I was still doped up and grief-stricken. Accepting that God has a plan doesn't always mean we have to like it. The constant grief persisted for a long time, and still comes in unexpected spurts, like when I realize my son would have started kindergarten this fall.

But that morning, July 19, 1979, I knew I did have a choice. Probably the most important choice I ever had to make regarding my life. I had to decide right then whether to remain mired in the events of the previous day, or to begin anew with my new found faith and move forward. I chose to move forward. Two years later we had a beautiful, healthy baby girl and I am trying to help her to learn to love God as I do. Hopefully she won't have to wait until she's 32 before she finds him to be her friend, as I did.

**Discussion.** I think this case and the open letter from this mother highlight many of the major issues we face in pediatrics, especially in the area of newborn medicine:

- Often the past relationship with a family, and the patient, is fairly limited. Obviously that's unfortunate in a human sense, but it also means we must make surrogate decisions in a void. Can I really know what a family wants? What impact will my decision have on the family unit, the marriage, other children? Do these parents act in the best interests of their child, or are they concerned primarily with their happiness and life? And, of course, should the family play any role in these decisions?
- If there is a place for the family, how does the physician balance the feelings and requests of the mother with those of the father? Do their religious convictions play a role in the plan?
- Can the physician, during a period of emergency care, become the surrogate decision maker if he or she is not always acting to preserve life?
- Must everyone partake of all that modern medicine can offer? Or put another way, is treatment always the best option?

# The Physician's Experience:

## Patients in Intensive Care

ANDREW B. LEIBOWITZ, M.D.

THE MAJORITY of ethical decisions are made by the most inexperienced physicians, at the worst times of day, and under the worst conditions possible. The average harried housestaff officer has little time—most often no time—to consider the intellectual, ethical, or moral ramifications of the life-and-death decisions he or she makes daily. Case 1 illustrates this point. Cases 2 and 3 are typical intensive care unit dilemmas.

Case 1. I remember vividly being a junior anesthesiology resident called to the medical intensive care unit to intubate a 29-year-old woman with respiratory insufficiency after vaginal delivery complicated by severe bleeding requiring hysterectomy. On my arrival, she looked me directly in the eye and emphatically stated she did not want to be intubated and realized she would probably die as a result of this decision. It was 2:30 AM, not exactly the best time to discuss life and death issues with the patient, her family, or other physicians. It was clear from her course that she was deteriorating rapidly and would succumb within a matter of hours, if not minutes, without intubation. Yet at this moment she was completely and totally coherent. What should I do? I heavily sedated her with Ketamine and intubated her against her wishes. The entire event, momentary indecision and all, took less than ten minutes. Because of the medication she had no recollection of the event. She required mechanical ventilatory support for several days and eventually returned home to care for her infant.

Case 2. A 67-year-old man with a long history of chronic obstructive airways disease and diabetes underwent a gastrectomy for gastric carcinoma. In the operating room it was discovered that he had metastatic disease. His postoperative course in the intensive care unit was complicated by respiratory failure, acute renal failure, and a stroke. It was the physicians' opinion that his long-term prognosis was nil and that even a short-term recovery was highly unlikely. His family felt differently and insisted on maximum supportive measures, including dialysis, tracheostomy, and—in the event of cardiac arrest—cardiopulmonary resuscitation. The physicians in charge of the case felt these measures were inappropriate, but in view of the family's strong desire for these therapies and their familiarity with the hospital and medical system, the patient was kept in the unit and dialyzed. When the patient's condition did not significantly improve after three days of aggressive treatment, these therapies were gradually reduced or eliminated and the patient was discharged, supported by a ventilator, to the floor. The family begrudgingly accepted this decision and the patient died two days after discharge from the unit.

Case 3. A 72-year-old man, a Jehovah's Witness, uneventfully underwent a coronary artery bypass and grafting. Prior to the procedure it was well established that he did not wish to receive any blood products even if failure to administer them resulted in his death. Two weeks postoperatively he suffered a major upper gastrointestinal hemorrhage, losing over half his blood volume; his hematocrit dropped to 12%. Again he refused blood-product administration. He was admitted to the intensive care unit and was treated aggressively with mechanical ventilation, paralysis, hypothermia (temperature of 90°F) and erythropoietin. He was supported in this manner for seven days. On the eighth day in the unit he experienced a myocardial infarction and died. There was only minor disagreement over honoring the patient's religious wishes, which directly caused his death.

---

Adapted from the author's presentation at the New York Regional Conference on Medical Ethics at the Mount Sinai Medical Center on March 9, 1990. From the Department of Anesthesiology, Mount Sinai Medical Center, New York. Address reprint requests to the author, Department of Anesthesiology, Box 1062, Mount Sinai Medical Center, One Gustave L. Levy Place, New York, NY 10029.

# The Physician's Experience:

## Elderly Patients

Diane E. Meier, M.D.

**Case Report.** The patient was an 81-year-old white woman when I began caring for her in the geriatric clinic in 1984. She was an alert, sharp-witted woman of fully intact mental capacity. Among her multiple medical problems were chronic obstructive pulmonary disease and bronchiectasis with recurrent pneumonia, congestive heart failure, chronic abdominal pain of unclear etiology, progressive vision loss from senile macular degeneration, chronic renal insufficiency, decreased hearing, and anxiety. Over the ensuing five years she suffered from progressive shortness of breath with exertion, such that by 1988 she was only able to walk about at home and required a wheelchair for all ventures out of doors. In the last two years of her life she was homebound except for her visits to the clinic. These activity restrictions were compounded by her worsening blindness, which left her unable to read or watch television.

She emigrated to New York City from Rumania in her twenties with her daughter. She had left her first husband several years earlier; the rest of her family was left behind in Rumania also, and many died in World War II. She was an accomplished seamstress and soon found work and a new husband, a widower with five little kids. She raised these children and continued working. Her second husband died when she was in her 60s but she kept working until she was in her late 70s. She was renowned for her baking and often cooked for family events. She lived alone in an apartment in Manhattan until about 10 years ago, when she acquired a 24-hour home health aide, Frances, who lived with her five days a week and became very close to her.

Over the last 10 years of her life the patient reported to me that her relationships with her daughter and stepchildren were unsatisfactory to her—she felt that they didn't visit or call enough and that they were ungrateful for her efforts during their childhood. She felt that the stepchildren resented her and that her own daughter was "too busy." She frequently reported that her current life was "not worth living" and repeatedly asked me for a prescription or a "pill," as she put it, that would allow her to end her life. Throughout this time she had no suicide plan, was sleeping and eating well, and had no manifestations of major depression. A trial of antidepressants failed, due to drug side effects. I always responded to her requests for a pill that would end her life by explaining that I couldn't do that, but that I would try to keep her as comfortable and symptom-free as I could.

However, when I asked her if she wished to see a living will or to assign durable power of attorney for health care to someone close to her in order to assure that her life was not prolonged when death was near, she refused to even look over such papers, stating that I knew what she wanted and that she didn't want to sign any paper. We had similar discussions many times over the last several years.

In the last 3 years of her life she was repeatedly hospitalized for recurrent pneumonias, hemoptysis, and gastrointestinal bleeding. She consistently refused all but the most noninvasive diagnostic and therapeutic measures but was willing to accept whatever treatments held promise of making her more comfortable. In the last year of her life she lost 30 lb (from her maximum

---

Adapted from the author's presentation at the New York Regional Conference on Medical Ethics at the Mount Sinai Medical Center on March 9, 1990. From the Department of Geriatrics, Mount Sinai Medical Center, New York. Address reprint requests to the author, Department of Geriatrics, Box 1070, Mount Sinai Medical Center, One Gustave L. Levy Place, New York, NY 10029.

of 100 lb) in spite of fairly good caloric intake. No evidence of neoplasm was found in admittedly limited diagnostic studies, and the weight loss was attributed to progression of her severe underlying lung disease. She developed pressure ulcers, in spite of superb home care, in the last months of her life and was admitted three times in 2 months for recurrent pneumonia. Efforts to establish home hospice services were unsuccessful, both because she was unwilling to sign the required DNR forms and was, in any case, unable to stay home when she became short of breath, because she preferred the relative security of the hospital.

I had little or no contact with her children until her final admission, when I notified the closest children, her daughter and youngest stepdaughter, that I felt she was unlikely to leave the hospital this time. Her dyspnea and hypotension improved slightly with 14 days of intravenous antibiotics and bronchodilators but recurred as soon as the antibiotics were stopped. Throughout the several-week period of this hospitalization, her mental status waxed and waned. She always recognized me and Frances and her daughters, but often thought she was at home or out on the corner when asked if she knew where she was. When I asked her if she wanted us to continue the intravenous antibiotics and medications, she varied in her responses, which included "I know you'll do the best you can for me, Dr. Meier" and "that's in God's hands." On one occasion she clearly stated she had "had enough" and that her time was coming. At that point she told her daughters she'd be joining their father in five days.

The family consistently pleaded with me to stop prolonging the agony and let their mother die. I was hesitant to do that, both because the patient's most recent statements reflected ambivalence about wanting to stay alive (recall "I know you'll do the best you can for me, Dr. Meier" and "that's in God's hands"), and because I was uncertain precisely whose best interests the family had in mind in their advice to me.

At this point I was concerned about several issues. First was the nature and quality of the relationship she had with her children. She was a difficult person, and it was clear that her children had painful and conflicted feelings about her. Were they really capable of exercising true substituted judgement? They reported that she had always been an extremely independent person who always did things for others and prided herself on her autonomy and strength. They felt that her progressive illness and disabilities over recent years had greatly affected her, causing depression, anger and a sort of helplessness. This was my experience with her also. Finally, they told me that she had said goodbye to each of them and had openly talked about some past disagreements and painful events, seeming to want to get these matters "off her chest." She did seem more peaceful as these meetings with her children were completed, began to sleep most of the time, and seemed less agitated and less short of breath.

The second matter I was concerned about was how to interpret her seemingly ambivalent feelings about life support. On the one hand she consistently repeated a wish for her life to be over and refused many diagnostic procedures with comments such as, "If I have cancer, I have cancer." On the other hand she was equally consistent in her refusal to codify these feelings by signing a DNR form or a living will. In fact each time I approached her with these concepts I felt almost as if I was assaulting her, as if she felt abandoned by me, as if I were breaking some important covenant with her. It was clear that she did *not* feel that these specific medical care decisions belonged with her and that she was very suspicious of any papers or forms that had legal or other implications. Did this reflect ambivalence about continuing wishes for life support and could the imminence of death have caused her to change her mind and did she now wish to stay alive as long as possible? Or was this simply a cultural or age-cohort aversion to signing forms or taking personal responsibility for health decisions? I wasn't sure, and her variable responses to queries about further antibiotics in the last weeks of her life only compounded my confusion.

After many long conferences with the two daughters, it became clearer to me that they did care a lot for their mother and had probably been more available to her over the years than she had reported to me. They felt that their mother's consistently stated desire to die, given the constricted quality of her life, was the truest reflection of her lifelong values of independence and productivity. Frances, the home attendant, who was really the closest friend the patient had, concurred.

I realized that regardless of what we did, her time was short because of the severity and end stage of her lung disease. Her ability to communicate was so curtailed by the illness that the *goals* of prolonging her life were increasingly unclear. In order to buy her a few more weeks, or at most months, of life she would require tube feedings and intravenous medications. The burdens of this approach began to seem to me to outweigh the potential benefits.

At the last I realized that both in terms of medical prognosis and respect for the woman's autonomy, further treatment was probably wrong. I recognized that she depended on *me* to make these decisions on her behalf, and I also recognized that I was having trouble letting her go. We did not reinstitute antibiotics, and she died in her sleep on the day she predicted she would, three days after intravenous antibiotics were stopped.

**Discussion.** I think this case sheds doubt on the usually prescribed nostrums for resolving ethical dilemmas in decisions to forgo life-sustaining therapy. First, it was difficult to rely on the patient's advance directives, both because she was inconsistent in her statements over time, particularly at the end of her life, and because she was unable to formalize her wishes in any way, a problem I have often encountered with older adult patients. They are frequently concerned that a signed living will or DNR order may be translated into medical abandonment.

My experience with some housestaff and their reactions to patients with no code orders suggests that indeed there is confusion about just what a DNR order means. It is often overinclusively interpreted and used as a justification to deny many other forms of care as well. In this era of escalating efforts at cost containment, I think we must take care to protect our patients from the denial of appropriate treatment, a process that is often hard to stop in the context of a DNR order or a living will.

The fact that advance directives may not be durable, that is that patients change their minds often, and often decline to put directives in writing, should not be surprising. Providing such advance directives requires patients to think in detail about what they want their deaths to be like, not an easy or pleasant task for most of us. Advance directives also require patients to clearly and coherently prioritize their life values, also not something most of us are given to; to think about how these values may be threatened under various scenarios; to understand the treatment options under these various scenarios; and to comprehend the implications of agreeing to or refusing treatment for each identified life value. Given these barriers, it is easy to see why advance directives are not commonly made in a formal way by many patients.

Second, the potential surrogate decision makers, her daughter and stepdaughter, and me, were handicapped as proxies. Her children had a lifelong history of conflict and ambivalence with her. Unquestionably her demanding nature was a burden to them. She herself repeatedly expressed anger and bitterness toward them to me, and would not identify either of them as proxies or give them durable power of attorney for health care decisions. Had I been an intern on the medical floors I would have known nothing of these strains in their relationships and would probably have accepted the daughters' substituted judgments without hesitating. But my longstanding relationship with the patient precluded this.

As her physician, I was somewhat protected from her prickly personality and was perhaps in a better position to try to identify the controlling values of her life and make a true substituted judgment for her. But try as I might to use what I knew of her so-called values history, her own ambivalence and frequent changes of mind made it almost impossible for me to step in and decide how she would have decided for herself, which is presumably what a surrogate decision maker is supposed to be doing. Several published studies from the University of Washington in Seattle and the University of Chicago, and unpublished data from our geriatric clinic, found a poor degree of agreement in medical decisions between patients of intact mental capacity and their potential proxies, who in all of these studies were close family members and physicians. In fact, the substituted judgments of these surrogate decision makers were no better than chance at predicting the patients' wishes for or against treatment in various scenarios.

In the end, then, I had to rely on medical prognosis and "best interest" criteria, a benefit-burdens analysis, in deciding what to do for the woman. In the end, neither her spoken advance directives nor the substituted judgments of surrogates were really helpful to me in arriving at that decision. I believe that this is often the case in decisions to forgo life-sustaining therapy and that more critical examination of these approaches is urgently needed.

# Surrogate Decision Making for Incompetent Adults:

## An Ethical Framework

Dan W. Brock, Ph.D.

My role in this discussion is to add a philosophical perspective to the consideration of surrogate decision making. In the medical ethics literature a substantial, though not universal, consensus has developed that health care decision making should be shared between the physician and the competent patient. Typically, in medical practice, the physician brings his or her knowledge, training, expertise, and experience to the diagnosis of the patient's condition and provides the prognoses associated with different treatment alternatives, including the alternative of no treatment. The patient brings his or her own aims and values to the decision-making process and ideally evaluates the alternatives, with their particular mixes of benefits and risks (1).

What values underlie a commitment to shared decision making, and what ends are we seeking to promote with it? The first, and most obvious, is the promotion and protection of the patient's well-being. Shared decision making rests in part on the presumption that competent patients who have been suitably informed by their physicians about the treatment choice they face are generally, though of course not always, the best judges of what treatment will most promote their overall well-being (2). The other value is respecting the patient's self-determination. I mean by self-determination the interest of ordinary persons in making significant decisions about their own lives themselves and according to their own values. It is by exercising self-determination that we have significant control over our lives and take responsibility for our lives and the kind of person we become. These then are the values that guide shared decision making.

For incompetent patients we still want shared decision making, but then the patient lacks the capacity to participate and a surrogate must take the patient's place. My aim in this paper is, therefore, to sketch an ethical framework for surrogate decision making about medical treatment for incompetent adults and for thinking about the ethical issues that arise in that decision making (3). Surrogate decision making should seek to extend the ideals of health care decision making for competent adults, with suitable changes required by and reflecting the patient's incompetence. Before considering who should be a surrogate and how the surrogate should decide, we must consider which patients should have a surrogate to decide for them.

### Determining Incompetence

In health care, if the patient is competent, the patient is entitled to decide and to give or refuse informed consent to treatment. If the patient is incompetent, a surrogate must be selected to decide for the patient. The concept, standards, and determination of competence are complex; and there is space here to emphasize only a few important points about competence and incompetence (see especially 3, ch. 1). Adults are presumed to be competent unless and until found to be incompetent. In questionable or borderline cases, competence should be understood as decision-relative, that is, a patient may be competent to make one decision but not another. First, decisions can vary in the demands they make on a

---

Adapted from the author's presentation at the New York Regional Conference on Medical Ethics at the Mount Sinai Medical Center on March 9, 1990. From the Department of Philosophy, Brown University, Providence, RI. Address reprint requests to the author at the Department of Philosophy, Brown University, 54 College Street, Providence, RI 02912.

patient, for example, in the complexity of the information relevant to the choice. Second, from the effects of medications, disease, and other factors, patients can change over time in the capacities they bring to the decision-making process.

Three distinct capacities are needed for competence in treatment decision making: the capacity for understanding and communication; the capacity for reasoning and deliberation; the capacity to have and apply a set of values or conception of one's good. Though people possess these capacities in different degrees, it is important to recognize that the determination of competence is not a comparative judgment. Because the competence determination in health care (and in the law) is used to sort the patient into either the class of patients who are competent to decide for themselves or into the class of patients who must have a surrogate to decide for them, it must be recognized as a threshold determination. The crucial question about competence in borderline cases then is how defective or impaired a person's decision making must be to warrant a determination of incompetence.

Two central values are at stake when a patient is judged competent or incompetent, and so two kinds of dangers should be balanced in that determination. One value is protecting the patient from the harmful consequences of his or her choice when the patient's decision making is seriously impaired. The other value is respecting the patient's interest in deciding for him or herself. The dangers to be considered are failing adequately to protect the patient from the harmful consequences of a seriously impaired choice, which must be balanced against failing to permit the patient to decide for him or herself when sufficiently able to do so. There is no unique objectively correct balancing of these two values and dangers. Instead, the proper balancing is inherently an ethically controversial choice.

**Process as the Standard.** The evaluation of a patient's competence should address and evaluate the *process* of the patient's decision making; the standard should not be an outcome standard that simply looks to the content of the patient's choice. Given the values at stake in the competence determination, it follows that the standard for competence should vary according to the consequences for the patient's well-being of accepting his or her choice. The standard should vary along a continuum from high (when the choice appears to be seriously in conflict with the patient's well-being) to moderate (when the patient's choice appears to be comparable to other alternatives in its effect on the patient's well-being) to low (when the choice will clearly best serve the patient's well-being).

One controversial consequence of this account of the competence determination is that a patient might be competent to consent to a particular treatment, but not to refuse it, and vice versa. This follows from two facts. First, the process of reasoning to be evaluated will inevitably be different if it leads to a different choice. Second, the effects on one of the values to be balanced (the patient's well-being—if the patient's choice is accepted) can be radically different depending on whether the patient has consented to or refused the recommended treatment. Treatment refusal may reasonably trigger an evaluation of the patient's competence, though it should also trigger a revaluation by the physician both of the treatment recommendation and of the communication of that recommendation to the patient; some studies have shown that the most common cause of treatment refusals is failure of communication between physician and patient, and most refusals are consequently withdrawn when the recommendation is better explained (4).

The critical question to answer is, Does the patient's choice sufficiently accord with the patient's own underlying and enduring aims and values for it to be accepted and honored, even if others, including the physician, may think it not the best choice?

### Selecting a Surrogate

Assume now that a patient has been found incompetent to make a particular treatment choice, so that a surrogate must act for the patient. Who should serve as surrogate? What standards should guide the surrogate's decision? If we are to respect the incompetent patient's wishes, then the surrogate should be the person the patient would have wanted to act as surrogate. In a number of states, by executing a Durable Power of Attorney for Health Care (DPOA), it is possible for a person, while competent, to legally designate a surrogate who will make health care decisions in the event of later incompetence. This document also allows a person to give instructions to the surrogate about one's wishes concerning treatment. Ethically, an oral designation by a person of a surrogate should have nearly the same weight as a formal DPOA. Competent patients should be encouraged to designate who will decide for them in the case of later incompetence. In fact, physicians have a responsibility to seek this information early in treatment and while patients are still competent, especially if a period of later incompetence is likely.

Often an incompetent patient will not have explicitly designated a surrogate. It is then reasonable and common to act on a presumption that a patient's close family member is the appropriate surrogate. (In most jurisdictions, a close family member lacks explicit legal authority to act as the patient's surrogate until formally appointed by a court as the guardian. States should consider adopting legislation like the recently enacted Health Care Decisions Act in the District of Columbia which formally authorizes a family member to act as surrogate for an incompetent patient without recourse to guardianship proceedings.) Only the most important considerations that support this presumption for a family member acting as surrogate can be discussed here. First, usually a family member will be the person the patient would have wanted to act as surrogate for him or her. Second, the family member will know the patient best and will be most concerned for the patient's welfare. It is important to underline that the claim is only that the *practice* of using a close family member as surrogate will result overall in better decisions for patients than any feasible alternative practice, such as appointing an attorney to act as the patient's guardian. Third, in our society the family is the central social and moral unit assigned responsibilities to care for its dependent members. Although dependent children are the most obvious example, dependent adults are another important instance of this responsibility. The family is also the main place in which most people pursue and realize the values of intimacy and privacy. Both to fulfill these responsibilities and to realize these values, the family must be accorded significant, though not unlimited, freedom from external oversight, intrusion, and control.

These grounds for presuming a close family member to be an incompetent patient's surrogate do not imply that such a surrogate must always make the optimal choice. On the other hand they also make it clear that family members' authority as surrogate decision makers is limited. When the grounds do not hold—for example, when there is no close relation between patient and family member, or when there is a clear conflict of interest between patient and family member—the presumption for family members as surrogates can be rebutted. In such cases, an incompetent patient's physician can have a positive responsibility not to allow the family member to act as surrogate.

Sometimes no family member is available to serve as surrogate. Most reasons that support a family member serving as surrogate will then also support using a close friend who is available and willing to serve. When no family member or friend is available, institutional flexibility is desirable because of the wide range of different decisions that must be made. Institutions only need a settled and public policy insuring that decision making in such cases does not become paralyzed from lack of a natural surrogate. For example, a hospital might have a policy for a specified range of decisions (such as decisions to forgo life-sustaining treatment or resuscitation) requiring that the attending physician's proposed decision be referred to the chief of service who could review the decision and take any further steps deemed appropriate (perhaps referral to an ethics committee or to the courts).

## How Surrogates Should Decide

**Advance Directives.** What standards should a surrogate employ in deciding for an incompetent patient? Three ordered principles can guide decision making: advance directive; substituted judgment; best interest. These are ordered principles in the sense that the surrogate should employ the first if possible; if not it, then the second if possible; and if neither the first or second can be used, then the third.

The advance directive principle tells the surrogate that if there is a valid advance directive from the patient, there is a strong presumption that it should be followed. Formal advance directives take two principal forms, living wills and durable powers of attorney for health care. Living wills are given legal force in approximately 40 states; DPOAs at present have legal force in far fewer jurisdictions.

Several common features limit the usefulness of living wills. I shall mention two. First, they are usually formulated in vague terms for describing both the patient conditions ("terminally ill") and the treatments ("extraordinary measures," "aggressive treatment"). Because they are executed well in advance of the decision to be made, this is to some extent inevitable and means that others must interpret how they should now apply. Second, probably in response to worries about potential abuses, enabling statutes in many states place limitations on the circumstances in which living wills can either be executed (for example, the person must already have been diagnosed as terminally ill with death imminent) or applied (for example, the decision cannot cover nutrition and hydration). When persons follow these restrictions and limitations, in many circumstances their living wills will fail to apply.

Since living wills are rarely brought into court for enforcement, but function primarily as a means of informing others about the patient's wishes, restrictive formulations may be inadvisable.

Except for those who have no surrogate, DPOAs are the more desirable form of advance directive for most persons. First, they allow a person to give more detailed instructions to the surrogate about wishes regarding treatment, though it is important to avoid letting greater detail narrow the application of the instructions. Second, they address the issue of later interpreting the patient's instructions by allowing the patient to designate the interpreter.

Although there is a strong presumption that a valid advance directive should be followed, there are reasons why advance directives should not have the same degree of binding force as the contemporaneous decision of a competent patient. The decision of a competent patient can be made attending to the full and detailed context, whereas advance directives must inevitably be formulated before the precise context of future decisions is known. This means that occasionally a treatment decision will arise in circumstances radically different from those imagined by the patient when executing the advance directive; in such cases, following the letter of the directive may be contrary to following its spirit. Moreover, concern over whether the patient would have changed his or her mind about treatment does not arise for a competent patient. Finally, a decision by a competent patient that appears significantly contrary to his or her interests usually will be challenged by the care givers, thereby testing understanding and resolve to an extent not possible with advance directives. Despite these limitations, there are good reasons to accord a strong presumption to any patient's advance directive.

**Substituted Judgment.** At present, and probably for the forseeable future, most patients do not have advance directives. Then the substituted judgment principle should be followed. This principle tells the surrogate to attempt to decide as the patient would have decided, if competent, in the circumstances that now obtain. In effect, the principle directs the surrogate to use his or her knowledge of the patient and the patient's aims and values to infer what the patient's choice would have been.

Physicians have an important responsibility in helping surrogates to understand their role in applying substituted judgment. Rather than asking, "What do you now want us to do for your mother?" the physician should say, "You, of course, knew your mother better than I did, so help us decide together what she would have wanted done for her now." Confirmation that asking the right question matters in the choices surrogates make can be found in Tomlinson (5). This substituted judgment approach is helpful for arriving at decisions which are more in accordance with the patient's wishes. The approach also has the practical advantage of making the psychological and emotional burdens easier for surrogates to carry and facilitating their effective participation in decision making.

"Do everything" is nearly always an inadequate and unhelpful answer. Surrogates should be assured that appropriate care will always be given, including all care needed to maintain the patient's comfort and dignity. But what care is appropriate will often change with changes in the patient's condition and prognosis, and surrogates must be helped to understand that all possible care is not automatically appropriate care.

Two features of substituted-judgment decision making should be explicitly noted. First, the substituted-judgment model will let surrogates take account of how the interests of others will be affected by the decision, to the extent there is evidence that the patient would have weighed those interests. Second, it will allow surrogates to assess the patient's quality of life, both at present and as it will be affected by treatment decisions, according to the patient's own values. For consideration of forgoing life-sustaining treatment only a limited judgment of quality of life is relevant: Is the best anticipated quality of life with life-sustaining treatment sufficiently poor that the patient would have judged it to be worse than no further life at all? No judgments about social worth or the social value of the patient are warranted under substituted judgment.

**Best Interest.** When no information is available about what this particular patient would have wanted in the situation at hand, the "best interests" principle directs the surrogate to select the alternative that furthers the patient's best interests. This, in effect, amounts to asking how most reasonable persons would decide in these circumstances, an approach which is justified by the absence of any information about how this patient differs from others. Treatment choices based on best interests can be especially difficult when reasonable persons can and do disagree about the choice. It is, therefore, important, where possible, to avoid having to appeal to best interests by determining patients' wishes about future treatment options when patients are still competent.

**Ordering These Three Principles.** Strict ordering of these guidance principles would be an oversimplification. Evidence bearing on the patient's wishes concerning the decision at hand, whether from an advance directive or from the surrogate's knowledge of the patient, is not either fully determinate and decisive on the one hand, or completely absent on the other. Instead, such evidence ranges along a broad continuum in how strongly it supports a particular choice. In all cases physicians and surrogates should seek confidence that the choice made is reasonably in accord with the patient's wishes or interests. The better the evidence about what this particular patient would have wanted, the more one can rely on it. The less the information and evidence about what this patient would have wanted, the more others must reason in terms of what most persons would want.

What constitutes adequate evidence about the patient's wishes has been an issue in several recent court cases, most notably the O'Connor case in New York and the Cruzan case in Missouri, recently upheld by the U.S. Supreme Court (6). In the New York case the court imposed an extremely high standard of evidence for forgoing life-sustaining treatment by surrogate order, requiring clear and convincing proof that the patient had made a settled commitment, while competent, to reject the particular form of treatment under circumstances such as those now obtaining. Where there is any significant doubt, the court reasoned, the decision must be on the side of preserving life. This decision set a very difficult standard in New York for patients and their surrogates to satisfy, and thereby establishes a strong presumption in favor of extending the lives of incompetent patients with life-sustaining treatment. In Cruzan, the U.S. Supreme Court upheld the right of the State of Missouri to impose this same strong presumption, though without endorsing the wisdom of doing so.

Such a presumption undervalues patients' interest in self-determination and fails to recognize adequately the extent to which patients' well-being is determined by their own aims and values. But is it not reasonable, the Court might ask, always to err on the side of preserving life when there is any doubt about the patient's wishes? Several decades ago, when medicine only rarely had the capacity to extend life in circumstances where doing so would have been unwanted and no benefit to patients, such a policy would have been reasonable. Medicine, however, has vastly enlarged its capacities to extend life. Patients' lives can now often be extended when they would not want this done, and, as a result, New York's and Missouri's strong presumption in favor of extending life, when there is any significant doubt about the patient's wishes, can no longer be justified. Though the U.S. Supreme Court found no constitutional bar to the clear and convincing evidence standard, good reasons remain for other states not to adopt it.

## Conclusion

I want to conclude with a plea for "preventive ethics" aimed at reducing the necessity to resort to surrogate decision making. This can only be accomplished by persons, while still competent, talking with their physicians and families about their treatment wishes should they become seriously ill. Such preventive measures are especially appropriate for persons with chronic, progressive diseases in which both a possible period of incompetence and the nature of later treatment decisions likely to arise are relatively predictable. Physicians have an important role in encouraging their patients to reflect on their wishes and to make those wishes known to those who are likely to be involved in their treatment decisions. The need for surrogate decision making in health care will never be eliminated, nor can all difficult decisions be avoided. But it should be possible greatly to reduce the number of cases in which physicians and surrogates must decide about an incompetent patient's care in the absence of knowledge of the patient's wishes when that information could have been obtained earlier had it been sought.

### References

1. Some respects in which this division of labor is oversimplified are explored in Brock DW. Facts and values in the physician/patient relationship. In: Veatch R, Pellegrino E, eds. Ethics, trust, and the professions. Washington, DC: Georgetown University Press, 1991.
2. For a discussion of different kinds of cases in which competent patients make irrational choices, and the responsibilities of their physicians in such circumstances, see Brock DW, Wartman SA. When competent patients make irrational choices. N Engl J Med 1990; 322 (May 31):1595-1599.
3. I draw freely here on prior published work I have done on this topic, some of it collaborative work with Allen Buchanan. See especially Buchanan AE, Brock DW. Deciding for others: the ethics of surrogate decisionmaking. Cambridge, England: Cambridge University Press, 1989.
4. Applebaum PS, Roth LS. Treatment refusal in the medical hospital. In: President's Commission for the Study of Ethical Problems in Medicine and Biomedical and Behavioral Research. Making health care decision: the ethical and legal implications of informed consent in the patient-practitioner relationship, vol. 2 Appendices. Washington DC: U.S. Government Printing Office, 1982.
5. Tom Tomlinson, et al. An empirical study of proxy consent for elderly person. Gerontologist 1990; 30(1):54–60.
6. In re O'Connor, No. 312 (NY Court of Appeals, October 14, 1988); Cruzan v. Director, Missouri Department of Health, 110 Sct 2841 (1990).

# Foster Parents as Surrogates for Infants and Young Children

Jonathan D. Moreno, Ph.D.

A formidable literature has developed on the rights and responsibilities of parents as medical decision makers, concentrated on the limits of authority of parents who are in general legally empowered to make decisions on behalf of young children. The limits of parental authority are usually defined in terms of the state's interest and society's obligation to protect children whose best interests are clearly in jeopardy (1).

By contrast, little attention has been given to the ethical implications of the situation of very young children in foster care who need medical attention, particularly for chronic illness. When the parents of infants in foster care are themselves too ill or uninvolved to make decisions about their child's medical care, who should be empowered to give consent? Should foster parents make medical decisions for the children in their care? Are there any issues about which foster parents should not decide? Who should decide when foster parents cannot?

The epidemic associated with the human immunodeficiency virus (HIV) is now forcing these neglected questions to our attention. As the number of children at risk for HIV increases, along with the frequency of parental inability to take care of them or even make decisions about their treatment, these questions take on increasing urgency. The ethical issues raised are especially complex because they appear in the overlapping territory of the foster care system and pediatrics. Each area is troubled by its own difficult problems of determining a child's best interests, respecting parental authority, and meeting societal obligations. In combination these problems can seem still more daunting.

In this paper I limit my attention to problems that specifically concern infants and other very young children in foster care. My comments and proposals apply to any infant whose condition requires regular or repeated contact with the health care system. Because AIDS in children raises especially pointed issues in this area, and because AIDS in particular has increased the burden on social services for children in some urban areas, I take the population of young children in foster care and at risk for HIV as my specific example later in the paper.

**Is Care Compromised by the Process of Consent?** It is notoriously difficult to assess the adequacy of health care for any population, partly because any such assessments involve counterfactual claims about what health status "would have been" had care been different. Quality assessment in health care generally is a notoriously contentious area. In particular, the adequacy of health care for young children at risk for HIV and in the foster care system has not been the object of a great deal of study. Further, because AIDS in children is still fairly new clinically, standards of care are not well-defined even in those parts of the country with substantial experience in HIV in children.

Child health care workers have suggested to me that nonemergent treatment of very young children in foster care may be delayed owing to difficulties in obtaining consent. Although attempts to investigate these claims have led only

---

A version of this paper was presented at a meeting of the Hastings Center project on "Caring for Infants and Young Children with or At Risk for AIDS: Ethical Issues in Diagnosis and Treatment," February 17, 1989. The preparation of this paper was partially supported by the George Gund Foundation.

From the Department of Pediatrics and Medicine, SUNY Health Science Center at Brooklyn. Address reprint requests to the author, who is Professor of Pediatrics and of Medicine (Bioethics) and Director, Division of Humanities in Medicine, at SUNY Health Science Center at Brooklyn, 450 Clarkson Avenue, Box 116, Brooklyn, NY 11203-2098.

to anecdotal material, I report these anecdotes, emphasizing lack of persuasive documentation.

Thus, for example, child welfare administrators describe instances in which nonemergent treatment for foster children has been delayed by the need to obtain the permission of a responsible representative of an agency, while a foster parent is standing by. Although there is no reason to believe that foster children are overtly injured by these treatment delays, if this claim is accurate they surely undergo some needless discomfort. Further, child care workers describe vastly different jurisdictional arrangements for obtaining consent for medical treatment. These inconsistencies could raise serious questions of inequities among children who are not relevantly different from one another.

Besides the adverse effects of delayed treatment for acute conditions, which are difficult to assess, health care providers claim to experience delays in permission to implement nonroutine diagnostic tests, such as CT scans. If the results of a test prove to have clinical significance, if for example a test discloses a tumor, delay in initiation of treatment could have irreversible consequences, particularly for a very young patient.

In light of these impressionistic and unsystematic reports, it might well be argued that my project in this paper suffers from lack of a definite problem. I can only say that the child care professionals with whom I have spoken are sufficiently distressed that it has seemed to me to be worth taking their concerns at face value. They seemed also to agree that, even if the dimensions of the problem are smaller than they believe them to be, the problems will almost surely increase as burdens on the foster care system grow in the near future.

**Partial Parenting.** Foster parents are increasingly called on to care for children with complex medical needs other than those posed by AIDS. Managing ventilators, suctioning, reinserting tubes, determining when dosages of medication should be altered, and recognizing emergencies have become necessary skills for many foster parents. After a few months as providers of this detailed care, the foster parents may be those most familiar with the typical responses of a chronically ill child, information that could be crucial but may not be readily available even to a health care professional. Yet when treatment decisions must be made and consent forms must be signed, the foster parent is rarely thought to be the individual with clear authority. Those who do have the authority—either social workers, other agency officials, or family court judges—have often had less contact with the child than the foster parents, or even none at all, and are understandably leery of taking this responsibility. This is one respect in which the foster parent is discouraged from developing any deep emotional relationship with the child in care, while also being deprived of the status and authority that are the basis for "parental tolerance, endurance, and devotion" (2).

If, as Goldstein and his colleagues assert, the preferred situation for a child approximates the emotional involvement that adults have with any wanted children, the foster parent role should present the opportunity for equivalent satisfaction in an adult-child relationship. Presumably an important element in the satisfactions these relationships provide is the adult's opportunity to play a role in the important decisions that shape a young life. Without a degree of control, the authors note, an adult's concern for a very young child is likely to be grounded only in that natural affection and concern that most human beings feel for the youngest members of the species, but for older children even that basic source of engagement is not dependable.

Of course one may dispute the authors' apparent assumption that biological parents do in fact have satisfactory emotional involvement with their children, particularly in light of rates of child abuse and neglect. One suspects that many foster parents cherish their foster children at least as much as many biological parents. My point here is only to illustrate the *theory* of foster parenting as found in a singularly important source, a theory that turns out to be wanting in its applicability to the current state of the foster care system.

In response, it may well be observed that foster parents have many other opportunities for deciding what is in a child's best interests in areas not involving health care, and that these are sufficient to establish an emotional connection. Further, one may argue that few of the health care decisions that must be made for a child are controversial and would not differ from one decision maker to another (anonymous reviewer's comments). But the first observation seems to underestimate the importance of health-care issues in the life of a child who is chronically ill, and the second perhaps gives short shrift to the sense of authority a parent normally feels in consenting to treatment, even though the substantive decision may not differ with the signatory of a form.

So far I have suggested that the theory of foster parenting recommends a degree of involvement with a child that may be imperfectly realized, given the realities of placement (3). What ethical conflicts could arise if foster parents in this system were given more authority to make medical decisions on behalf of children?

**Sorting Out the Moral Issues.** Biological parents are usually assumed to be concerned with their child's best interests, and thus are the presumptive decision makers for their children; a similar presumption about foster parents does not obtain. Why is this the case?

One reason is the temporary nature of foster placement; the expectation of a continuing relationship that ordinarily conditions a guardian's emotional involvement with a child is absent. Another is the explicitly financial nature of the placement contract, which creates opportunities for at least more direct conflicts of interest than is normally the case for biological or adoptive parents. A third reason is only apparently circular: foster parents cannot be legally responsible for their foster children because they are not the legally recognized parents; if they were legally responsible they would not be *foster* parents and, therefore, no unique issues would arise.

More plausible than the designation of foster parents as presumptive medical decision makers for the children in their care is the conclusion that in some circumstances they should be empowered to make some medical decisions. But if they are permitted a major role in deciding what a child's best interests are, what is to prevent them from interpreting those interests in terms of their own religious or philosophical beliefs, including some not within the mainstream of this society? Perhaps some procedural safeguards should be created. Are there some categories of decision that should not be made by foster parents, but rather by agencies or courts? Should there be routine procedures to review medical decisions by foster parents?

A related factor is the continuing authority of the biological parent who has lost custody of the child, especially when the foster parent has been given increased authority to make decisions, among them medical. Sufficient legal grounds for relieving the parent of the custodial role may not be sufficient morally to ignore the biological parent's preferences entirely. One can imagine a range of circumstances in which a parent's moral rights to participate in decision making for a child can be distinguished from his or her legal rights. Indeed, the moral right to full parental authority and the moral right merely to participate in major decisions about the life of one's child need to be more sharply distinguished.

For example, the parent who has a history of wanton abuse of a child may be said to have ceded moral as well as legal rights to make decisions for that child. However, the mother who immediately following childbirth is judged unable to care for her newborn due to her drug addiction and its associated lifestyle is a less obvious case. Do all parents who make poor choices about their lives give up their moral rights to participate in their children's lives? In particular, are all the rights of motherhood breached if a woman exposes her fetus to the consequences of her own disease? The problems associated with maternal-fetal conflicts have generated their own literature (4). Although these issues are beyond the scope of this paper, they do create specific moral questions for those involved in the foster care of a child born under such circumstances.

**Ethical Decision Making for the Newborn.** Many of the perplexing ethical issues that concern medical decision making for infants arise in the neonatal intensive care nursery (5). Although the analytic framework that has emerged in the ethics of newborn care was not designed specifically to deal with problems arising in the child welfare system, that framework has implications for this topic, particularly in respect to the continuing involvement of biological parents.

Ordinarily, parents have the right to decide about treatment. But, at least in what might be called the "received view" in medical ethics, that capacity is constrained in two extreme conditions. In the first, the parents prefer an option contrary to the best interests of the child in that it implies undertreatment, as in the case of an infant with Down's syndrome who needs an operation to permit normal ingestion, but whose parents decline the procedure. In the second, the parents prefer an option that implies overtreatment, for it would only prolong the dying process and the child's suffering, as is the case of infants who are ventilated though stricken with massive cerebral bleeding incompatible with survival. In still other cases, parents are paralyzed by shock and sorrow, in effect forcing others to become actively engaged in decision making for their child.

When parents make decisions that health care professionals believe are clearly incompatible with the best interests of a newborn, or if the parents' inability to decide itself constitutes a risk to the child, options may be pursued that transfer legal guardianship to another party. In general, however, caregivers are loathe to take such extreme measures unless all other alternatives have been exhausted. Recognizing that the fragile connection between the parent and the very young child is important for both and can easily be damaged, doctors, nurses, and social workers prefer to negotiate a path through the crisis.

Thus an application of the intensive care nursery model to decision making in foster care suggests, first, that the priority and involvement of the biological mother should, with rare exceptions, be sought and preserved. The only legitimate exceptions to this standard would be those instances in which the interests of the child are clearly and imminently at hazard, threatening consequences for the child even graver than the risk of permanent loss of the maternal bond.

Moreover, as suggested above, the biological parent without legal custody is not necessarily bereft of a legal or moral right to participate in making decisions about a child. At the least, the parent who has lost normal parental rights at childbirth is entitled to general information about the child's placement and, if the child becomes ill, about his or her condition. Similarly, parents who no longer have custody and who are themselves terminally ill are entitled to identify relatives whom they would prefer as foster parents for their child.

Finally, the intensive care nursery model indicates that no parent, whether biological or foster, is entitled to make decisions for a child that clearly do more harm than good, even if those decisions stem from deeply held religious beliefs. Included in this limitation on authority would also be decisions that increased a child's dependency or limited his or her desirability as a candidate for future placement, without corresponding benefit to the child.

**Categories of Protected Decision Making.** Are there some categories of medical decision making from which foster parents should be excluded? Under prevailing regulations in virtually all jurisdictions, for children who are in the custody of a commissioner of social services, foster parents can consent to routine care, but may not consent to any intervention regarded as invasive.

Yet this general restriction by degree of invasiveness seems unreasonable. Rather than referral to a social service agency for all prospective nonemergency interventions—a process that further burdens the system and often involves serious delays—some interventions are, albeit "invasive," clearly in the child's best interests and present little or no risk (X-rays to diagnose pneumonia, sutures for lacerations, a cast for a broken limb). Therefore, many low-risk interventions may be candidates for foster parents' consent.

But medical treatment that presents significant risks to children should probably be excluded from any plan to broaden the authority of foster parents. Procedures that require the use of general anesthesia or that involve withholding or withdrawing life-sustaining therapy are paradigmatic. Thus foster parents may want to know the seropositivity status of an infant at risk for HIV. But a positive test can stigmatize a child and should, therefore, be considered a significant social risk that cannot be taken by the foster parents alone.

Considering areas in which the rights of the socially displaced or handicapped have been violated in the past, one category that should probably be excluded is that of reproductive decisions. This matter will have particular poignancy if and when HIV-infected children survive puberty. Moreover, determinations in this area lend themselves to confusion between the interests of children and those who care for them: it is much easier to render a sexually active teenager infertile than to influence and monitor his or her behavior. Judgments about which form of birth control should be adopted, whether an abortion should be performed or an adolescent sterilized should be made through a process that allows the greatest possible accountability and consideration of presumptive reproductive rights.

**Conditional Authority for Foster Parents.** Certain reforms in the system of foster child care could improve the early medical care of children as well as relieve the system of some degree of burden in responding to proposed interventions. Treatments or tests that may qualify are those that are neither emergent nor routine but are clearly medically indicated and involve virtually no medical, psychological, or social risk to the child.

However, as has already been indicated, not all foster parents are in a good position to take this responsibility, and not all biological parents are entirely uninvolved or unresponsive to entreaties about a child. Therefore, for foster parents to qualify for conditional authority to give consent for the kinds of interventions that have been described, at least two conditions should hold: first, the natural family must be nonresponsive; and second, on assessment there must be very favorable bonding between the child and the foster parents. Even with this authority, foster parents should be required to report every instance of its use immediately to the child's caseworker, who would in turn report it to the biological parents. Further, all foster parents who are assessed as qualified for this conditional authority should pass through a rigorous additional training period that brings them into closer collegial contact with the agency, which remains responsible to the social services commissioner for the well-being of the child.

There are a number of difficulties with this proposal. First, rather few foster parents may qualify for this level of authority. Second, for a time at least, agencies may find themselves engaged in more rather than less work with the foster parents. Third, the arrangement may encourage some foster parents to have exaggerated notions about the extent of their authority. Fourth, if it succeeds it may confusedly be thought of as a way to compensate agencies for a shortage of professional caseworkers.

But these problems with a system of conditional authority for foster parents seem manageable and are surely less compelling than the prospect that some children will benefit from more efficient medical attention and more involved foster parents. Although other foster parents might resent those who are in this sort of program, they might also see it as a mark of status and wish to improve their own skills in order to qualify.

**The Informed "Silent" Parent.** Consider finally the rarer case of an HIV-infected child in voluntary temporary foster placement, a situation that can occur if, for example, the mother with HIV decided that she cannot take care of her baby but does not want to give up her parental role entirely while she is alive. How can foster parents achieve the "important relationship" with such a child while also helping to maintain the biological parent's rightful connection with the child? The intensive care nursery model again presents a clue. In that setting caregivers often act as de facto agents for parents, helping them to come to terms with a decision or, if the parents prefer, effectively making the decision for them.

When foster care is being arranged, a diagnostic intake study is routinely performed by an agency. This study assesses the physical, social, and psychological conditions that will affect a particular placement. Part of the study may also include agreements between the agency , biological parents, and foster parents. At this point the biological parent could be given the opportunity to assign decision-making authority to the foster parents through the agency, subject to revocation at any time. During the period of assigned authority the biological parent would continue to be provided with all information about the child's condition and progress, as well as what treatment alternatives were available and were chosen.

The advantages of this arrangement accrue to all parties. The child is more likely to receive needed medical care in a timely manner. The foster parents are more likely to feel involved with the child because they are empowered to take steps on his or her behalf. And the biological parent retains her rights to information about her child's condition and the ultimate authority to end the agreement and regain an active voice.

The disadvantages of this "silent parent" arrangement include the possibility that the biological and foster parents will be somewhat confused about their exact roles in relation to the child. But this is endemic to foster care and if anything may be ameliorated with the proposed arrangement. Another, more serious problem is that the biological parent could feel coerced into accepting a placement in which she virtually loses control over her child. One way to prevent this would be the establishment of a rule that prohibits agencies or foster parents, or both, from insisting on parental transfer of medical decision-making authority or on its continuation as a condition of placement. "Silent parenting" would, therefore, begin as an option and would continue to be an option thoughout foster placement. The "silent parenting" option could, however, be revoked at any time without loss of the foster home.

Health care providers may understandably be reluctant to do a test or procedure with only foster parents' consent. So a legal mechanism must be created that reflects the "silent parent" arrangement. This can be done in a document, executed by an attorney retained by the child welfare agency, that enables the biological parent to retain custody of the child while appointing the foster parents guardians of the *person* of the child unless and until one of the parties, the biological or foster parents or the agency, revokes the agreement.

The mother who has lost custody of a child immediately following its birth but who has not, in fact, engaged in child abuse should also be a candidate for this "silent parent" arrangement. The only difference would be that she would not have the opportunity to take decision-making authority on herself again. For although normal parent-child relations have not been initiated in these cases, the day may come when a relationship is established—though it may not come until the child has reached legal majority—and society should encourage whatever sense of identification by a parent with a child is possible insofar as it does not conflict with the child's best interests.

### Acknowledgments

The author expresses his gratitude to Mary Banach, Jeff Blustein, and an anonymous reviewer for their very helpful comments on earlier drafts, as well as several very helpful informants in the child welfare system.

### References

1. Buchanan A, Brock D. Deciding for others: the ethics of surrogate decision-making. New York: Cambridge University Press, 1989.
2. Goldstein J, Freud A, Solnit J. Beyond the best interests of the child. New York: Free Press, 1973:25.
3. Kline D, Overstreet MF. Foster care of children: nurture and treatment. New York: Columbia University Press, 1972:218.
4. Legal interventions during pregnancy. Court-ordered medical treatments and legal penalties for potentially harmful behavior by pregnant women. JAMA 1990; 264:2663–70.
5. Caplan A, Cohen C, eds. Imperiled newborns. Hastings Center Report 1987; 17.

# Panel Discussion

## The Ethics of Surrogate Decision Making

Moderators: Daniel A. Moros, M.D., and Rosamond Rhodes, Ph.D.

**Rosamond Rhodes:** I'd like to invite Judge Kaye to begin with any questions or comments that she may have for the presenters.

**Hon. Judith Kaye:** When we completed our legal presentation we had a crisp package of rules and principles which seemed to dissolve when we were confronted by a series of difficult cases. Then the philosophers refocused our attention on rules. All of us, physicians, philosophers, advocates, and jurist, reflect the different laboratories in which we work. Here at Mount Sinai you live with individual cases and attempt to apply principles. In the law, and particularly where I sit, we are making abstract rules that hopefully will be sensibly applied across the board. I must resist my tremendous desire to respond to some of the comments on the O'Connor case—especially those from Prof. Brock—because since the O'Connor decision was issued, it has been sensibly applied. I don't know how many of you are familiar with the Terry Coove case, a decision that came after the O'Connor case. There the trial judge applied those "impossible" criteria of the O'Connor decision and signed an order permitting the extubation of an 86-year-old woman. When the ink was barely dry on the order the woman opened her eyes. Someone asked her if this was what she would want. Almost her very first words were, "Well, that's a very difficult decision to make."

**Jonathan Moreno:** When I am confronted with a case, I find that the differences in disciplinary perspective between me, a nonphysician, and the physicians and nurses and social workers tend to wither away. A physician sometimes presents a case and says, "See, this guy got up and thanked us, and you tell me not to be paternalistic." If we were to look at cases prospectively, and try to pay attention to our thought processes, we'd probably find a lot of agreement.

**Rhodes:** Could we look at the issue of paternalism in a somewhat different way? As I understand, physicians are bound to follow the Hippocratic Oath or, for some, the Oath of the College of Surgeons, to act for the good of the patients. That gives them a clear duty, much like the duties of a parent to the child, or the teacher to the student, or the government to the population. In the history of philosophy we have, right at the beginning of the *Republic,* the story of a fellow with his spear. The fellow says to his friend something like, "Hold this spear for me for awhile." The friend returns crazed. Perhaps he'd gone into the baths, seen his wife in the arms of someone else, and now wants to do them both in with the spear. He says to his friend, "Give me back my spear!" The question is, does justice require the friend to return the spear? Everybody says, No—because it's not good for the other person. And we have this sense, very strongly, that our duty is to act for the good of the other.

Then we have, from the law and from philosophers, another sense, that we must respect autonomy. But at the same time, we think, as moral beings, we have to act for the good of another. In terms of the law, or in terms of medical practice, do you think you are obliged to act over the expressed directions of another for their good?

**Ian R. Holzman:** I am convinced, from the families I have seen, that a significant minority of parents do not act in the best interests of their child in the newborn nursery. That is, they clearly are acting in their own best interests.

---

Adapted from the concluding panel discussion at the New York Regional Conference on Medical Ethics at the Mount Sinai Medical Center on March 9, 1990. Address reprint requests to Rosamond Rhodes, Coordinator of Medical Ethics Teaching, Box 1193, Mount Sinai Medical Center, One Gustave L. Levy Place, New York, NY 10029.

Now, they don't put it in those words, but this sort of circumstance is the basis of the Baby Doe regulations. Probably the only thing I like about the Baby Doe regulations is that I can invoke them fairly early on in discussions, when the first words out of the parents' mouths are, "This baby isn't going to be normal and, you know, we really can't deal with that. We have another child at home, and can't you stop giving this kind of intensive care?" Although I'm willing to answer for myself and say, "I don't want to do that," I usually get around to saying, "Well, we can't do that in New York State because of the Baby Doe regulations."

I don't think we can assume parents always act in the best interests of the child.

**Moreno:** I think the question is, when a parent walks in, what do we assume? I think the presumption ought to be that this parent is going to make decisions in the best interest of the child. It might not be true one out of ten times, or one out of six times, or one out of five times, but the other four or five or nine times, it's going to be true. To assume otherwise deprives a parent of the opportunity to make a decision about that child for whom the parent, more than anyone else in the world, cares. In a multiethnic, multiracial environment, there are powerful differences between the way people look and act and talk, differences that make it difficult to have confidence in our presumptions.

**Fenella Rouse:** When you're talking about best-interest decisions, either for children or adults, there may be four or five or six decisions that are all within an acceptable range of best interests. It's not clear in each case that there is one right decision and one bad one. That's very important. In the O'Connor case, for instance, if you think it was better for her to be alive than dead (assuming that we don't know what her wishes really were), was it better for her to be alive for a year with tube feeding (which is what happened), or for four or five months with spoon-feeding, in which case she would still have had interaction with people? Both alternatives could arguably be in her best interest.

Doctors here, in their individual discussions, talk about a lot of decisions that were probably illegal. I have always maintained that the law is not workable in every case. But, as Dr. Holzman has indicated, that law is often invoked when a doctor wants to use it. According to my clients, health-care workers say, "You know, I'm awfully sorry, the law says so-and-so." What they mean is, "I'm not comfortable with you making this decision." This attitude toward law is not fair.

**Kaye:** I want to exercise the prerogative extended to me initially to put a question to the panel, one that follows up on Dr. Holzman's comment, as well as Fenella Rouse's last comment. How much does the law actually mean to you? How much does it determine decisions that you're making every day? Dr. Meier, in the kind of decisions that you make and that you described, how much weight do you give to the law?

**Diane E. Meier:** That's a tough question. I guess I would say not too much, in that I don't allow it to really determine my decisions. But it has an impact, and we certainly do think about it and talk about it. For example, in the case that I presented, I was concerned because I was documenting on the chart what my patient was saying every day, and she was saying, "I know you'll do what's right for me, Dr. Meier." That could have been interpreted in a court of law as her saying, "Keep me going." I didn't. In that case I decided to stop treatment, on the basis of best interest. I think that when we hear about cases like the O'Connor case, we despair at our efforts to obtain advance directives, because none of them are specific to this disease, this treatment, under these circumstances. Nobody can anticipate that kind of detail about the preterminal phases of an illness. So, when one tries to practice in New York State, one feels discouraged about the state of the law, at least the case law.

However, in the case I presented I was attempting to demonstrate the practical problems within this hierarchy of decision making endorsed by the judge, legal speakers, and philosophers. In this particular case I attempted desperately to use a standard of advance directive and a standard of surrogate decision making, and I guess I would have to agree, after listening to you, that I did make use of these standards. I didn't make a decision independent of the patient's repeatedly stated wishes to die. I certainly took that into account. I took into account the daughters' opinions after I had spoken with them for a long time. But ultimately the decision was left to me, and I think I made it on the basis of "best interest," although I didn't specifically use one standard in preference to the others. In practice, things are not that clear.

**Dan W. Brock:** I started out saying that I was going to give you three ordered principles, and then after talking about them, noted that, in fact, if you don't have completely decisive evidence concerning a patient's wishes, substituted judgment becomes the only relevant principle, and you can only reason in terms of best interest. One must seek assurance that the decision is reasonably in accord with the wishes of the patient, to the extent those are known, and the best interest of the patient. This means that to the extent that one has less than decisive evidence about a patient's wishes, different kinds of reasoning will be involved in different cases. You should try saying to yourself, I'm uncertain about what this patient wants. How can I match that up with what most people seem to want in these circumstances and with what I know about the patient?

I want to say something more general about this contrast between the way physicians have been talking about the cases, and the way the lawyers and philosophers have been at least taken to be talking. Although philosophers and lawyers like to order things, at this conference we are not seeking a framework which could be mechanically applied and somehow transform cases you rightly perceive to be difficult into cases that are easy to resolve. An ethical framework of the kind I was talking about is nothing more than an initial way to help order thinking about difficult cases. I suspect that the way we think about these cases isn't so different from the ways the physicians do.

**James J. Strain:** The complexity and circumstances of medical decision making are a problem that needs to be emphasized. For example, if you reason on the basis of best interests, you must confront the conflict between empathy and sympathy. If a comatose patient is admitted and I don't know the patient at all, I follow the principle of best interests. If I ask myself, "How would I feel if I were in that situation?" that's sympathy. But I really should call on empathy and ask, "If I were that person, in that situation, in their shoes, what would I do?" This is a different order of thinking.

I don't make decisions on the basis of how I might feel if that were my mother or my uncle. I make the decision on the basis of how that person feels, based on my intuition and whatever evidence I can gather about them. The decision might be entirely different if I use empathy rather than sympathy.

Empathy is not carefully considered in philosophical argument. I don't know how a judge can get to appreciate the extent to which empathy plays a role in a physician's day-to-day decision making. That data is never presented to a judge.

Another example: If my conscious communication expresses the belief that I don't want to live, and that thought is recorded and documented and the judge gets it, that's a certain kind of data. But maybe my actual position is that I don't want to live *if* I am going to be in pain, or I don't want to live *if* there is absolutely no chance of me either eating or swallowing again, or I don't want to live *if* I am never going to be able to hear Beethoven's Ninth again. What you really need to know is the meaning of the communication. What bothers me about the philosophical rules and precision you're demanding in the decision-making process is the issue of the psychological assessment. For the psychiatrist making the psychological assessment it demands that we take upon ourselves the responsibility for interpreting the meaning of the communication with empathy.

One last example, following Dr. Moreno's discussion: If you were an HIV-positive child in this city, a foster child, you can only get AZT if you are entered in a research protocol. [Henry Sacks is in the audience and has been faced with this problem in his research proposal.] The foster family, the biological family, and the Bureau of Child Welfare must all sign off on the case in order to get permission for a foster child to be entered into a protocol. The only way a foster child can get the medication is through a protocol. This is a serious impediment against the child in foster care getting AZT, the only treatment that one might get for AIDS.

Similarly, in order to protect prisoners on Rikers Island from abusive experimentation they cannot be AZT experimental subjects. Now these prisoners are petitioning to go onto the experimental drugs. It seems to me that we are on a fast track for rethinking a lot of paternalistically protective rules on a more sophisticated level.

**Brock:** I agree with most of what you said. I don't see why you think that the philosophers don't want to look below the surface of what a patient is saying to you. If you take substituted judgments seriously, it tells you to try to understand the ambiguous evidence at as deep a level as you can and examine behavior which may belie what the patient is explicitly saying. You may have to understand how things like depression can be distorting what the patient is now wanting, and that depression can lead to questions about the patient's competence, and so forth. I don't think anyone wants to embrace a notion of autonomy which views whatever the patient seems to be saying as automatically the deciding factor. The challenge is always to try to understand, in as deep a way as we can, what patients do want, and we must recognize there are going to be cases in which, despite our best efforts, reasonable people will still disagree.

**Rhodes:** I hear a sharper disagreement than you suggest. Your view of having substituted judgment or autonomy rule amounts to permitting patients to choose their own moral standards. That choice could be egoism; it could also be beneficence. A particular person could always choose to do what's least good for himself or herself and best for her or his family. But the doctor, or whoever it is, who acts for the good of the patient is always presuming that the patient's moral theory is egoism. This is the point that Dr. Moros has made. Adopting substituted judgment requires that you examine what Dr. Meier would call the person's value history, and insert that person's values, rather than the person's best interests, which would always be egoism. That sounds like a real distinction.

**Brock:** There does seem to be a conflict. Physicians take a Hippocratic Oath which commits them to the patient's well-being. On the other hand, public policy says physicians aren't empowered to do whatever *they* think is best for patients, and that's why we have a doctrine of informed consent.

**Rhodes:** What the government thinks and what the physician thinks can be two different things.

**Brock:** Most of the physicians I work with don't think that they should do what they think is best for patients without regard for what those patients want. So I don't think physicians think that informed consent is just an external intrusion of government into what they should be doing. My experience is that physicians don't necessarily feel uncomfortable when faced with a patient who wants to accept something less than what might be best for the patient, perhaps because the patient cares about a family member. Physicians don't try to talk that patient out of that kind of altruism.

**Rhodes:** What you say pertains to the competent patient, but does it hold for the patients for whom you'd want a surrogate to decide?

**Brock:** A surrogate could say, "You know, it looks like doing A would be best for this patient, but I think the patient probably would have wanted to do B because he would have cared more for the effects on these other people." When you've got a surrogate making that kind of judgment, I think we appropriately ask for a relatively high standard of evidence that the patient would have sacrificed his or her apparent interests for the well-being of another. But when a surrogate has good enough evidence, then I think physicians appropriately accept that kind of decision. The more the sacrifice of the patient's interests for the benefit of someone else, the higher the standard of evidence should be.

**Strain:** I think that's why we're having conferences like this.

**Meier:** There are thresholds for paternalism and for responses to patients who refuse what you recommend. If the patient refuses this particular antibiotic, because it gives them a stomach ache, okay, you'll prescribe a different antibiotic. If, on the other hand, a patient refuses a simple procedure (such as placing a nasogastric tube that you think might be necessary for two weeks to get them through a postoperative period) one would, I think appropriately, escalate one's level of paternalistic intervention and persuasion because one feels that the patient's capacity to make a decision of that magnitude simply isn't sufficient. And so I think you are constantly adjusting your level of intervention and paternalism in proportion to the magnitude of the decision being made and the capacity of the patient to respond.

**Rhodes:** But this response also suggests that your ultimate view is restoring autonomy to the patient.

**Meier:** Absolutely.

**Brock:** It's important to add that it's not just from the physician's point of view that this gets done. As potential patients, all of us would want our physicians to push harder when more is at stake.

**Meier:** It should be looked on as a safety net that most rational people would want beneath them in the event of loss of capacity.

**Rhodes:** You are suggesting a new model for surrogate decision making, the model of the rational person deciding. This would be an alternative to deciding according to what the patient would have wanted and also an alternative to deciding according to the patient's best interest.

**Moros:** It's hard to envision what the patient intended when writing an advance directive. What does it means to you on that day or in ten years? The only directives I have any comfort in following are those from patients who I feel shouldn't be resuscitated and who it would be medically inappropriate to resuscitate.

**Leibowitz:** I have a similar problem. Every single medical study on resuscitation shows that many patients who medically require resuscitation will not survive the resuscitation. In fact, most of the patients for whom a physician would actually consider signing a DNR order have less than 1% chance of surviving an arrest. In the intensive-care unit, where I work as an intensivist, we routinely do not code patients who have no hope for survival.

One day a hospital administrator came to me and said, "What do you mean you're not going to code Mr. C in bed 8?"

I said, "He has a condition that is not remediable. CPR was designed to treat cardiac arrhythmias. That's not his problem. That's not the indicated medical therapy." She responded that the hospital legal position was that everybody gets coded even if it's not an indicated procedure when there is no signed DNR form.

**Holzman:** This emphasis on resuscitation is irrational and misplaced. It is irrational to resuscitate someone who has no chance of living. Many, many physicians in our circumstances don't believe that everybody needs to be continually resuscitated, DNR orders notwithstanding. Resuscitation is a treatment to bring someone back to life, not to torture someone continually because there's a law that says you have to have these papers signed.

**Rhodes:** We started with Judge Kaye, and Judge Kaye will have our last comment.

**Kaye:** We have focussed a bit on the inadequacies and limitations of the law. But the law and cases such as these have done a tremendous amount to raise the public consciousness about all of these issues. I don't think I've read a record on any of these cases in which the person hasn't said, I don't want to be like Karen Ann Quinlan. And of course, right now we have the case in the Supreme Court of a young woman, Nancy Cruzan, who has been in a persistent vegetative state for seven years and could go on for decades more. [Cruzan has since died.] We've had television programs, we've had articles in *The New York Times*, a tremendous article in the *Sunday Times*. I think people are confronting their mortality, and in a sense, the kinds of issues we're all talking about here. I'll claim that as a benefit of the legal system. Surely that's in the interests of all of us.

# *2. Drug Use: License vs. Prohibition*

# History of Drug Use and Drug Control

Daniel M. Fox

---

My role in initiating this symposium is to provide some background about the history of drug abuse and drug control. I want to explore two distinct but related questions. First, what is the history of addiction and the efforts to control the use of narcotics in the United States? Second, what can this history tell us that may be useful in analyzing drug abuse in the late 20th century? I begin by focusing on three broad themes in the history of drug abuse and drug control: the ubiquitousness of drug use, the integration of drug use within cultures, and what I call the limited repertoire of policy.

**Ubiquitousness of Drug Use.** To begin with, one must recognize that intoxicants have been used at some time in the history of most recorded cultures. The purposes of intoxicants are different in different cultures and over time within the same culture. What is ubiquitous is the use of intoxicating substances, rather than the use to which they are put. Also, the array of intoxicants changes over time within cultures and within particular subcultures. The universality of drug use, therefore, challenges the current description of our situation as a new epidemic of substance abuse.

**Cultural Integration of Drug Use.** If the use of intoxicants is a common feature of human culture over time, then the history of intoxicants and their use will be integrated into the history of many cultures. To understand how people have eaten, what they have eaten, what they have drunk, and why they ate and drank what they did brings one to the heart of understanding a culture. Intoxicants appear as an important element of work, kinship, family relations, religion, and, certainly, sex. Intoxicants have been linked with gender, gender roles, gender behavior, and also with sexual relationships. Intoxicants and war have been even more intimately related. Furthermore, changes in what is defined as medical care invariably alter patterns and uses of intoxicants.

Some specific examples from United States history illustrate the cultural integration of intoxicant use. Wine, beer, and whiskey are important in the behavior patterns and food habits of the various cultures to which Americans belong. When the Erie Canal was being built, a ration of a quart a day of gin was handed out by the contractors to the people who dug the canal, as part of their compensation for working. Whiskey was a part of the laborer's culture, and it was also a stimulant.

Another example is the hypodermic needle; it was introduced in the United States during the 19th century as a means for injecting morphine. Only later was it adopted for use in the control of infectious diseases. Historians and historical epidemiologists concur that most narcotic addiction in the 19th century was incidental to the practice of medicine. From the scant records on addicts prior to the second decade of the 20th century, it appears that most became addicted as the result of medical treatment. Since access to medicine has always been strongly correlated with class status, it should not be a surprise that addiction to morphine and other medicinal drugs tended to be found in the middle and upper ends of the class scale.

We do not know the extent of serious addiction problems among physicians in the 19th century. The euphemism "impaired physician" was not yet in use; however, we do have illustrative stories.

One story, in particular, has been central to the history of American academic medicine. It concerns William Stewart Halstead, the founding

---

Adapted from the author's presentation at the New York Regional Conference on Medical Ethics at the Mount Sinai Medical Center on March 3, 1989. From the Milbank Memorial Fund.

professor of surgery at the Johns Hopkins School of Medicine. He became addicted to cocaine in 1886, while a young academic physician in New York City. He spent most of the next two years in hospitals trying to shed this dependence. The medical literature typically reports that Halstead was cured, with only occasional lapses throughout his career, a magnificent career central to the development of American surgery. It turns out that his hospitalization replaced his cocaine addiction with morphine addiction: the morphine addiction was lifelong. It also turns out that the behavioral patterns on the surgical service at Hopkins—an organization imitated by most surgical departments in the United States—include the frequent absences of the professor and the high fees he charged, all suggesting a pattern consistent with Halstead's ongoing addiction.

The United States' acquisition of an overseas empire between 1898 and 1904 certainly influenced American attitudes toward drug use. Our involvement with the Philippines, in particular, made narcotics use and transportation a major problem for the United States. Further, massive immigration in the latter half of the 19th century introduced into the population people from cultures with different ways of using and managing narcotics.

Through the 19th century and well into the 20th, various intoxicating substances, not just alcohol but also opium derivatives and cocaine, were sold over the counter. They were used without any of the moral stigma attached to drinking alcohol or being addicted to drugs. Industrial investment in the promotion of intoxicant use was large. The pharmaceutical industry, which at times in American history has been indistinguishable from the patent medicine industry, has had a major impact on drug use and regulation. It appears that big drug houses in the early 20th century were unable to maintain adequate security at their facilities; through theft they became a major source of black-market opiate sales. Some historians actually argue that New York City's place as a center of illicit drugs was not so much a result of immigration or internal migration as it was the product of poor security at the drug factories then located in New York City.

I need not say too much about the alcohol or tobacco industries' interest in maintaining social acceptance of addiction to their products. Although I have never seen any evidence of United States government complicity in fostering tobacco addiction to maintain the tax yield, in England some recently uncovered files reveal that, in the 1950s, at least some individuals in the Treasury sought to interfere with the Health Department's support for publication of Sir Richard Doll's studies on the carcinogenic effects of tobacco.

A word about popular notions of the history of Coca Cola: the word "coca" used in the name suggested the drug cocaine, and this drug, then believed to be nonaddicting, could have been supposed to be a substitute for alcoholic beverages. Did the company include cocaine in the drink, and then take it out? In the late 1890s, the institutionalized racism (school and residential segregation, the poll tax) that marred American history in the 20th century was rapidly emerging. Hate-mongering toward Afro-Americans included stories claiming that black men, crazed from drinking Coca Cola or using other drugs, were committing rape. Historians have found no documented cases of cocaine-induced rape in the American South during the 1890s. The stories, however, continued to circulate that the company changed its secret formula and removed cocaine from its beverage.

Another cultural point about intoxicants in United States history is change in the perception of their effects over time. Cocaine was not initially viewed as addictive, and even as recently as the early 1980s some serious scientists made that claim. Alcohol has been a part of the nutritional patterns of many cultures and, until recently, the use of tobacco was fully accepted in our society. During the late 19th century, heroin was introduced from Germany as an acceptable pharmaceutical. The American view of the effects of these drugs has certainly shifted, and the demography and epidemiology of addiction has also changed. The population addicted to opium and its derivatives during the 19th century was drawn from the affluent section of society, particularly woman. The addict population today is dramatically different. Even the definition of pathology and our view of what is pathological has changed. Is it pathological to smoke a cigarette, a pipe, or a cigar? These changes make it difficult to claim that current drug use is unusual or new. Addiction is not epidemic, it is endemic.

**Limited Repertoire of Public Policy.** The repertoire of public policy to confront addiction is limited. The three classic policy tools are social control, tolerance, and therapy. Each is rarely used by itself; public policy typically combines all three approaches, making it difficult to evaluate the effectiveness of any effort. There are tremendous problems in gathering information on the demography and practices of drug use. Given these difficulties, we must question whether much of the information collected can count as data at all. Can it be used to inform rational public policies of social control or tolerance, or is the information merely rhetorical ammunition for moralism, ideology, and political agendas?

Looking back at the prohibition of alcohol (1920–1934), the most famous imposition of the three combined policies, the complexity of the problems is evident. Clearly, national prohibition, and the state-by-state prohibition that preceded it (in more than 30 states), was a measure of social control. Certainty the existence of an often unimpeded underground market in alcohol can be seen as a measure of social tolerance. (Like the market in health care, distribution of illegal alcohol was skewed toward those who could afford it.) During these years of prohibition, as a therapeutic consequence of the policy, the rate of cirrhosis of the liver was cut in half.

American popular thinking about drug use since the turn of the century as revealed in newspaper and magazine photographs seems to be that drug use occurs mainly among outsiders to our culture, is not a central feature of our social world, and is somebody else's problem. Characteristically, early in the century, the Chinese are the opium smokers and dealers. In the 1930s marijuana smokers are not quite underworld figures, but are pictorially set among the shadows of distant figures. Social control is celebrated before World War I in photos of burnings of captured opium. The dominant image of drug control throughout the century is not the therapeutic image, it is the enforcement image—often a heroic enforcer showing off drugs seized in raids, and thus showing the effectiveness of social control as a policy. In the current worldwide epidemic of HIV infection, one characteristic popular pictorial metaphor is the hypodermic needle—that marvelous medical invention for pushing morphine on the middle-class population.

# The Epidemiology of Cocaine Use in New York State

Blanche Frank, Ph.D., Don C. Des Jarlais, Ph.D., Rozanne Marel, Ph.D., James Schmeidler, Ph.D., and Michael Maranda, M.Phil.

Ethical and legal considerations about the problem of drug use and drug abuse are complex and involve numerous value and practical conflicts. Debate over the proper ethical and legal courses of action is often hampered by the poor quality of current data. Collecting accurate data on licit drug use is difficult, but collecting data on illicit drug use is extremely difficult. As our contribution to the present discussion of ethical concerns in drug use, we present the currently available data on the epidemiology of cocaine use in New York State. We have selected cocaine because, of all illicit drugs, it currently appears to pose the greatest problems for both public health and public safety. We examine in detail the use of cocaine in the state and the association of cocaine use with the AIDS epidemic in New York.

## Cocaine Use among Adults in New York State: Survey Results

Periodically the New York State Division of Substance Abuse Services sponsors surveys of households to assess the extent of illicit drug use, the patterns of use, and the consequences of use. The last two surveys were conducted in 1981 and 1986 (1, 2). In each survey, questionnaires were administered by telephone in an interview averaging 30 minutes. (The telephone has become the medium of surveying due to escalating costs of face-to-face interviewing and the difficulty of gaining entry into households.) A pilot telephone study in 1980 had shown the effectiveness of telephone surveying in epidemiologic studies of psychoactive prescription drugs used nonmedically and the use of illicit drugs (3).

The 1981 survey on illicit drug use was conducted by Chilton Research and included 3,487 respondents. The 1986 survey was conducted by Louis Harris and Associates, and included 6,364 respondents. In most respects, the surveys were comparable. The sampling strategies in both surveys used random digit dialing with the same geographic stratification. Basic interview questions were the same. The 1986 survey, however, asked about mode of administering cocaine (intranasal use, injection, "smoked"), whereas the 1981 survey did not. The 1986 survey included only respondents 18 years and older, whereas the 1981 survey included respondents 12 years and older. For purposes of comparison, the following analysis presents findings for those 18 years and older projected to the total adult population, 12.5 million in 1981 and 12.7 million in 1986.

Prevalence. Overall, the 1986 survey found that 9% of the population, or more than 1.1 million New York State residents, 18 years of age or older, have used cocaine in their lifetime (Table 1). In terms of illicit drug use, this number is second only to the number of those the survey found who have tried marijuana (3.6 million, or 28% of the adult population), but exceeds the number who have tried hallucinogens, such as LSD and mescaline (553,000, or 4%).

---

Adapted from the presentation by DCDJ at the New York Regional Conference on Medical Ethics at the Mount Sinai Medical Center on March 3, 1989.

From the New York State Division of Substance Abuse Services, New York, NY (BF, RM, MM), Beth Israel Medical Center, New York, NY (DCDJ), and Mount Sinai School of Medicine (CUNY), New York, NY (JS, formerly of New York State Division of Substance Abuse Services). Address reprint requests to Blanche Frank, Ph.D., New York State Division of Substance Abuse Services, 55 W. 125th Street, New York, NY 10027.

TABLE 1

*Cocaine Use among the New York State Population by Area of the State, 12,669,000 Residents, 18 Years of Age and Older, Spring 1986*

| Area | Never used n (%) | Ever used n (%) | Recent use (past 6 months) n (%) | Current use (past 30 days) n (%) |
|---|---|---|---|---|
| New York State | 11,504,000 (91) | 1,147,000 (9) | 499,000 (4) | 279,000 (2) |
| New York City | 4,597,000 (87) | 661,000 (13) | 290,000 (5) | 182,000 (3) |
| Suburban New York City* | 2,419,000 (91) | 233,000 (9) | 112,000 (4) | 51,000 (2) |
| Upstate urban counties† | 1,651,000 (92) | 134,000 (8) | 53,000 (3) | 23,000 (1) |
| Other counties‡ | 2,837,000 (96) | 118,000 (4) | 45,000 (2) | 23,000 (1) |

Numbers may not add up to state totals due to rounding.
* Suburban New York City includes the counties of Nassau, Suffolk, Westchester, and Rockland.
† Upstate urban counties include Albany, Erie, Monroe, and Onondaga.
‡ Other counties include New York State counties exclusive of those in New York City and the counties of Nassau, Suffolk, Westchester, Rockland, Albany, Erie, Monroe, and Onondaga.

Respondents were also asked about the most recent time they used cocaine—whether in the past six months (recent use) and the past 30 days (current use). Almost half a million adult residents (499,000, or 4%) reported the use of cocaine in the six months prior to the survey, and more than a quarter of a million (279,000, or 2%) used cocaine in the 30 days prior to the survey.

The populations of the several areas within the state vary in rates of cocaine use. As expected, New York City shows the highest rates of cocaine use in the adult population, 13% having ever used the drug at some time, 5% having used it in the six months preceding the survey, and 3% currently. Rates in the suburban counties around New York City are exactly at the state average; the next highest rates occur in the largest urban counties of upstate New York. The lowest rates are in the remaining counties of the state, 4% of the adult residents having used cocaine, 2% using recently, and 1% currently.

TABLE 2

*Trend in Cocaine Use among the New York State Population, 18 Years of Age and Older by Area of the State, 1981 and 1986*

| | 1981 n (%) | 1986 n (%) |
|---|---|---|
| Ever used | | |
| New York State | 702,000 (6) | 1,147,000 (9) |
| New York City | 368,000 (7) | 661,000 (13) |
| Suburban New York City* | 163,000 (6) | 233,000 (9) |
| Upstate urban counties* | 57,000 (3) | 134,000 (8) |
| Other counties* | 114,000 (4) | 118,000 (4) |
| Recent use (past 6 months) | | |
| New York State | 431,000 (3) | 499,000 (4) |
| New York City | 241,000 (5) | 290,000 (5) |
| Suburban New York City* | 99,000 (4) | 112,000 (4) |
| Upstate urban counties* | 29,000 (2) | 53,000 (3) |
| Other counties* | 62,000 (2) | 45,000 (2) |
| Current use (past 30 days) | | |
| New York State | 226,000 (2) | 279,000 (2) |
| New York City | 124,000 (2) | 182,000 (3) |
| Suburban New York City* | 51,000 (2) | 51,000 (2) |
| Upstate urban counties* | 12,000 (1) | 23,000 (1) |
| Other counties* | 39,000 (1) | 23,000 (1) |

Numbers may not add up to state totals due to rounding.
* Area definitions as in Table 1.

**Comparison of 1981 and 1986 Findings.** The change from 6% to 9% is a significant increase in experience with cocaine in New York State in the five years between 1981 and 1986 (Table 2). Although increases were found in many areas throughout the state, in the less urban counties in upstate New York essentially no change occurred in the rate of initial cocaine use between 1981 and 1986; however, the numbers of "ever in lifetime" users in the rural counties should not be taken as evidence that there were no new users of cocaine from 1981 to 1986. The 1986 number may have been reduced through failure to report by persons who had used the drug in the distant past and by migration of users (young adults) out of rural areas.

The comparison of rates of recent cocaine use is even more interesting because it shows essentially no change for the state as a whole or for the areas within the state. For instance, in 1981 3% and in 1986 4% of the state's residents had used cocaine during the previous month. Although many more New York State residents had tried cocaine by 1986, it seems as if an almost constant proportion of residents are using the drug at any one time. This finding suggests that many residents have made occasional use of cocaine and only about 2% actually use it at any time. Even at this rate, more than a quarter of a million New York State residents, or 279,000 adults aged 18 years and older, used cocaine during a recent month.

**Frequency of Use.** Although a similar proportion of New York State residents are using cocaine recently or currently, those using it in 1986 were more intensively engaged than the comparable population in 1981. In 1981 20% and in 1986 41% of current users had taken cocaine

TABLE 3

*Frequency of Cocaine Use in the Past Month among Current Cocaine Users in the New York State Population by Area of the State, 279,000 Current Cocaine Users, 18 Years of Age and Older, Spring 1986*

| | Current use (past 30 days) | Used once | Used 2–3 days | Used 4–9 days | Used 10 + days | Used, frequency unknown |
|---|---|---|---|---|---|---|
| Area† | *n* (%) | *n* (%) | *n* (%) | *n* (%) | *n* (%) | *n* (%) |
| New York State | 279,000 (100) | 112,000 (40) | 49,000 (18) | 66,000 (24) | 48,000 (17) | 4,000 (1) |
| New York City | 182,000 (100) | 69,000 (38) | 34,000 (19) | 40,000 (22) | 35,000 (19) | 4,000 (2) |
| Suburban New York City | 51,000 (100) | 18,000 (36) | 8,000 (15) | 17,000 (34) | 7,000 (15) | ** |
| Upstate urban counties | 23,000 (100) | 16,000 (67) | ** * | 2,000 (10) | 5,000 (22) | ** |
| Other counties | 23,000 (100) | 8,000 (36) | 7,000 (32) | 7,000 (30) | ** (2) | ** |

Numbers may not add up to state totals due to rounding.

* Less than 0.5%.

** Less than 500.

† Area definitions as in Table 1.

four or more days during the previous month. In 1981 8% and in 1986 17% of current users were using the drug 10 or more days during the previous month. This tendency to more frequent use among users occurred in each area of the state (Table 3). As expected, this tendency is most dramatic in New York City and its suburbs: in 1981 41% and in 1986 49% of current users used cocaine on four or more days of the previous month. Use weekly or more often occurs to a lesser extent in upstate urban counties (32%).

New Use. In both the 1981 and the 1986 survey, respondents were asked when they first used cocaine. The proportion of the population who started use during the year prior to each survey gives a comparative measure of the spread of cocaine use among adults, 18 years and older (Table 4). Interestingly, this rate of incidence remained remarkably stable—1% of the population initiated use in 1980–1981 and in 1985–1986. Rates in areas of the state also changed little. It is possible, however, that the initiation rates during the intervening years may have been higher.

## Patterns of Use

In the past few years, modes of administration other than sniffing and snorting have gained popularity. Smoking cocaine, which can result in severe dependency, has become especially popular. In addition, cocaine users are now using other substances, sometimes to enhance the effects of cocaine, sometimes to ameliorate its effects, and sometimes as part of a multiple-drug-use lifestyle. The 1986 survey inquired into modes of administration, use of other drugs, and drugs used in combination with cocaine.

**Route of Administration.** As expected, sniffing or snorting is the most popular way in which cocaine is used; 9% (1,089,000 residents) of the total population have engaged in it (Table 5). Still, a substantial proportion of the population of New York State have smoked or used cocaine in freebase form—2% or 289,000 residents. By 1986, 1% of residents had used crack and slightly less than 0.5% had injected cocaine. However, because of the heavy social stigmatization of illicit drug injection, compared to other forms of illicit drug use, underreporting of this behavior is a problem. We do not consider the reports of injection to represent the actual number of persons who have injected cocaine in New York.

When respondents were asked about their mode of cocaine use, they were certain about whether they had sniffed or snorted, injected, or smoked. When, however, those who indicated smoking or freebasing the drug were asked whether they had ever used crack, a projected 40,000 residents were uncertain. This result may be partially attributed to the fact that at the time of the survey, in the spring of 1986, crack was not widely known throughout the state, particularly in the upstate counties. Exactly what constitutes crack and whether or how it is distinguished, in the user's mind, from other forms of cocaine processed for smoking, remains unclear to this day. Consequently, the findings for crack use in this

TABLE 4

*Trend in New Cocaine Use among the New York State Population by Area of the State, 1981 and 1986**

| | 1981 | 1986 |
|---|---|---|
| Area† | *n* (%) | *n* (%) |
| New York State | 161,000 (1) | 166,000 (1) |
| New York City | 83,000 (2) | 79,000 (1) |
| Suburban New York City | 36,000 (1) | 28,000 (1) |
| Upstate urban counties | 11,000 (1) | 35,000 (2) |
| Other counties | 31,000 (1) | 25,000 (1) |

Numbers may not add up to state totals due to rounding.

* New cocaine use defined in 1981 survey as first use between January 1980 and spring 1981; defined in 1986 survey as first use between January 1985 and spring 1986.

† Area definitions as in Table 1.

TABLE 5

*Route of Administration of Cocaine among the New York State Population by Area of the State, 12,669,000 Residents, 18 Years of Age and Older, Spring 1986*

| Route of administration of cocaine† | New York State | New York City | Suburban New York City‡ | Upstate urban counties‡ | Other counties |
|---|---|---|---|---|---|
| | *n* (%) | *n* (%) | *n* (%) | *n* (%) | *n* (%) |
| Ever sniffed or snorted | 1,089,000 (9) | 613,000 (12) | 232,000 (9) | 130,000 (7) | 114,000 (4) |
| Ever injected | 41,000 * | 34,000 (1) | 5,000 * | 2,000 * | ** * |
| Ever smoked or freebased | 289,000 (2) | 151,000 (3) | 64,000 (3) | 39,000 (2) | 35,000 (1) |
| Ever used "crack" | 69,000 (1) | 50,000 (1) | 5,000 * | 9,000 * | 6,000 * |

Numbers may not add up to state totals due to rounding.

* Less than 0.5%.

** Less than 500.

† Categories of route of administration are not mutually exclusive; respondents may be counted in more than one.

‡ Area definitions as in Table 1.

TABLE 6

*Use of Other Drugs in the Six Months prior to the Survey among 499,000 Recent Users of Cocaine in New York State, by Area of the State*

| | New York State | New York City | Suburban New York City† | Upstate urban counties† | Other upstate counties† |
|---|---|---|---|---|---|
| | *n* (%) | *n* (%) | *n* (%) | *n* (%) | *n* (%) |
| **Recent use* of cocaine** | 499,000 (100) | 290,000 (100) | 112,000 (100) | 53,000 (100) | 45,000 (100) |
| Recent use* of other drugs | | | | | |
| Marijuana | 396,000 (79) | 226,000 (78) | 91,000 (81) | 41,000 (77) | 38,000 (84) |
| Other illegal drugs‡ | 83,000 (17) | 54,000 (19) | 6,000 (5) | 6,000 (12) | 17,000 (38) |
| Prescription drugs used nonmedically§ | 65,000 (13) | 35,000 (12) | 15,000 (13) | 10,000 (19) | 5,000 (11) |

Numbers may not add up to state totals due to rounding.

* Recent use: use in six months prior to survey.

† Area definitions as in Table 1.

‡ Other illegal drugs: one or more among heroin, illicit methadone, hallucinogens (e.g., LSD or PCP), and inhalants (e.g., amyl/butyl nitrite or glue or aerosol sprays).

§ Prescription drugs used nonmedically: use for purposes other than prescribed, in amounts greater than prescribed, and/or received from source other than personal prescription of one or more among tranquilizers (e.g., Valium and Librium), sedatives (e.g., Quaalude and Tuinal), stimulants (e.g., Preludin and Ritalin), analgesics (e.g., Darvon and Percodan), and cough medicine with codeine.

1986 survey probably undercount the number of users. Nevertheless, for analytic purposes and because of the addictive potential associated with smoking cocaine, a hierarchy was constructed with respect to using cocaine in this manner. Thus, of the 1,147,000 New York State residents aged 18 years and older who reported ever having used cocaine:

- 69,000 reported the use of crack and most probably the use of cocaine by other means;
- 220,000 who did not report the use of crack reported smoking or freebasing cocaine;
- 857,000 who neither reported using crack nor smoking or freebasing cocaine reported other means of use—mainly sniffing or snorting.

**Use of Other Drugs.** Use of additional drugs, either by themselves or in conjunction with cocaine (Table 6), is another indication of the seriousness of cocaine use. Of the 499,000 New York State residents who used cocaine in the six months prior to the survey, 79% had also used marijuana during that period, 17% used other illegal drugs (hallucinogens, narcotics, inhalants), and 13% used prescription drugs (tranquilizers, sedatives, stimulants) nonmedically. Almost every recent cocaine user also reported the use of alcohol. Users also took other substances along with cocaine, typically whatever was most available; 61% of recent cocaine users concurrently used alcohol and cocaine, and 56% concurrently used marijuana and cocaine.

Although in the suburbs the concomitant illegal drug is likely to be marijuana, upstate urban residents show a relatively strong tendency to also use psychoactive prescription drugs nonmedically, and residents of less-urban upstate counties show a relatively strong tendency to supplement their cocaine use with marijuana or other illegal drugs. The unmistakable conclusion of the data analysis is that cocaine users are also using other drugs.

Close analysis shows that these patterns exist within all of the areas of the state. In New

York City and its suburbs, cocaine users follow a wide array of substance use patterns.

### Cocaine Use and AIDS in New York

The acquired immunodeficiency syndrome (AIDS) has been considered the greatest public health challenge facing the nation. Although infection by human immunodeficiency virus (HIV), the causative agent for AIDS, is commonly associated with the injection of heroin, the injection of cocaine has been an important part of exposure to HIV in New York. HIV was first introduced into New York's intravenous drug use community sometime in the 1970s (for a more complete history of the HIV epidemic among such drug users in New York, see ref. 4). The number of HIV-exposed IV drug users stayed relatively low for several years, and then, during the late 1970s and early 1980s, the virus spread rapidly.

Since the mid 1980s, the HIV seroprevalence rate among IV drug users in Manhattan has stabilized at approximately 50%. This stabilization reflects some new HIV infections each year within the pool of IV drug users, balanced by the loss of HIV seropositives through death and also the entry of a number of new, seronegative persons into active IV drug use, offset by the loss of some seropositive individuals who discontinue IV drug use. Some AIDS risk-reduction behavior among current IV drug users has also contributed to the stabilization of HIV seroprevalence among IV drug users in New York.

The earliest blood samples showing HIV exposure among IV drug users come from a series of liver disease studies conducted between 1978 and 1983 (4). Recent analyses of the drug-injection behavior of the participants in these studies show that cocaine injection is a statistically significant independent predictor of exposure to HIV (5). Data collected from New York IV drug users between 1984 and 1988 also show that cocaine injection is related to HIV seroprevalence. (In these studies, cocaine injection and heroin injection were combined into a single variable of drug injection frequency; 6, 7). This correspondence holds both in studies of IV drug users who are not in treatment (8) and in treatment program studies. One study of new HIV infections which separated cocaine from heroin injection showed that cocaine injection was the factor most strongly associated with seroconversion (9, 10).

Clearly cocaine injection has played a significant part in the spread of HIV among IV drug users in New York City. There is also the possibility that noninjection cocaine use may become associated with sexual transmission of HIV. Cocaine can be smoked in the form of crack or freebase. The arousal effects of smoking cocaine are often interpreted as sexual feelings by males, leading to a desire to have sex while smoking or shortly after smoking crack. A sex-for-crack type of prostitution has arisen in the "crack houses" in New York and other large American cities. Ethnographic observations of these crack houses indicate that most of the sexual activity must be considered unsafe. Trading sex for crack has been associated with increases in syphilis in Philadelphia (11). Syphilis and other sexually transmitted diseases that cause open lesions can greatly increase the transmission of HIV. If IV drug users who have been exposed to HIV through sharing injection equipment have sex in crack houses, these locations could become the scene for heterosexual transmission of HIV. Thus, it is quite possible that cocaine use could develop a second important route in the transmission of HIV in New York.

### Discussion

As background material for discussions of the ethical issues and public policy approach to drug uses we chose to focus on the epidemiology of cocaine use. Widespread experience with the drug, as well as the complexity and severity of the public health problems associated with its use, justified this selection.

**Change as the Normal Condition.** The data presented could support a variety of policy and ethical positions. With this in mind we raise one caution for those developing public policy and ethical positions on cocaine use. Many discussions of ethical and legal concerns regarding drug use and drug abuse start with the assumption that the "present situation" will continue unless new policies are adopted. Epidemiologic studies show that, instead, one should view change as the normal condition. Over the last decade, most of the change in cocaine epidemiology has had dire consequences for public health in New York. AIDS has emerged as an unexpected and catastrophic consequence of injecting cocaine. Although the level of cocaine use in the general population remained relatively constant, there has been a clear shift to more intensive use among persons using cocaine. Predicting the short-term future is difficult. There are signs of possible improvement; drug injectors in New York have been changing their behavior to reduce the risk of developing AIDS (9). There are also signs that the problem could grow much worse if crack use increases or if crack-for-sex becomes an important mode of HIV transmission. In any case, with cocaine use, a larger population may be entering the pool of those at increased risk for contracting HIV infection.

We suggest that policy makers considering the complex ethical and legal questions involved in cocaine use avoid focusing on the design of a single "best policy." Rather, they should aim at designing a flexible system which would permit rapid public health responses to reinforce positive trends and to mitigate adverse trends in cocaine use. We need to think in terms of continually changing patterns of drug use rather than limiting ourselves to solving yesterday's problem.

## References

1. New York State Division of Substance Abuse Services. Preliminary report: drug use among New York State's household population, 1981.
2. Frank B, Marel R, Schmeidler J, Maranda M. Cocaine use among New York State residents. New York State Division of Substance Abuse Services, 1987.
3. Frank B. Telephone surveying for drug abuse: methodological issues and an application. In: Self-report methods of estimating drug use: meeting current challenges to validity. NIDA Research Monograph 57, 1985; 71:83.
4. Des Jarlais DC, Friedman SR, Novick D, et al. HIV-I infection among intravenous drug users in Manhattan. JAMA 1989; 261(7):1008–1012.
5. Novick DM, Trigg JL, Des Jarlais DC, et al. Drug abuse patterns and ethnicity in IVDA during the early years of the HIV epidemic. Presented at the Fifth International Conference on AIDS, Montreal, 1989.
6. Marmor M, Des Jarlais DC, Cohen H., et al. Risk factors for infection with human immunodeficiency virus among intravenous drug abusers in New York City. AIDS, 1987; 1:39–44.
7. Schoenbaum EE, Selwyn PA, Klein RS, Rogers MF, Freeman K, Friedland GH. Prevalence of and risk factors associated with HTLV-III/LAV antibodies among intravenous drug abusers in methadone programs in New York City. Presented at the 2nd International Conference on AIDS, Paris, 1986.
8. Friedman SR, Des Jarlais DC, Kleinman P, et al. Risk reduction among intravenous drug users in and out of treatment. Presented at the Fourth International Conference on AIDS, Stockholm, 1988.
9. Des Jarlais DC, Sotheran JL, Stoneburner R, et al. HIV-I associated with fatal infectious diseases other than AIDS among intravenous drug users. Presented at the Fourth International Conference on AIDS, Stockholm, 1988.
10. Ethnographic studies suggest that the frequency of cocaine injection may make "safer injection" difficult to practice. Persons injecting cocaine may do so as frequently as every five minutes. Under these circumstances, if injection equipment is being shared within a group, it is difficult to properly disinfect the equipment between each use, even if bleach is available for disinfection. Friedman SR, Sterk C, Sufian M, Des Jarlais DC. Effects of changes in drugs injected on drug using environments and on risk reduction among IV users. Presented at the Fifth International Conference on AIDS, Montreal, 1989.
11. Rolfs R, Goldberg M, Alexander ER, Sharrar R. Drug related behavior and syphilis in Philadelphia—sex for drugs. Am J Epidemiol 1988; 128:898.

# Paternalism and Drug Abuse

TZIPORAH KASACHKOFF

IN THIS PAPER I am setting myself the limited objective of outlining what I believe to be some of the important issues concerning paternalism and drug abuse. First, I shall set out my understanding of the term "paternalism" and the way in which it is relevant to the topic of this conference.

By "paternalism" we generally mean actions or (in the case of governments or institutions) regulations, policies, or laws whose purpose it is to prevent an individual from harming him or herself or—more positively and broadly—whose purpose it is to promote that person's welfare. Generally speaking, then, paternalism is the name for actions or other ways of affecting others where the motive for one's conduct is the desire to help another. It is, in short, being our brothers' and sisters' keepers.

What makes paternalism an issue of contention is that much of what we can do to promote another's welfare, or to keep her from harm's way, can be done only by interfering—sometimes without welcome—in the life of the person we wish to benefit. Indeed, it is the intrusion into another's life, as much as the altruistic motive for doing so, which has come to characterize that conduct we call "paternalistic." Because of this, "paternalistic actions" stand in need of some defense: since we value the right of individuals to lead their own lives, and to conduct their lives as they rather than as others see fit, any interference with another's life, even for the purpose of furthering or protecting that person's interests, demands some justification.

The question about paternalism raised by the topic of drug use is whether there is any justification for paternalistically interfering with those who may be termed "drug abusers." To what extent, if any, are we morally permitted to intrude into the lives of drug abusers in order to further their interests or to protect them from harm? This question will demand different answers from the government, family, friends, parents, pastors, physicians, and children of drug abusers, for whether it is morally right (or morally obligatory) to paternalistically interfere with a drug user will depend, to a large extent, on whether we are the friend or the spouse, the physician or the pastor, the parent or the child, of the drug abuser.

**Paternalistic Legislation on Drug Use.** The question we want to address here is whether the government and the law are justified—and if so, to what extent and in what ways—in intruding into the lives of citizens for the purpose of protecting them from the dangers of drug abuse. This is the question of the legitimacy of "legal paternalism" with respect to drug abusers.

Legal paternalism is not, in general, unfamiliar to us: our government has made it illegal to leave school and to buy alcohol before a certain age; to drive cars without seatbelts; to ride motorcycles without helmets; to buy certain drugs without prescriptions; to sell cyclamates; to refuse to contribute to a social security retirement plan; and to lend money at usurious rates—to mention only a few of our paternalistic laws. These laws are designed, and defended—at least insofar as they are paternalistic—as protections of ourselves from actions that result from our own folly, ignorance, negligence, laziness—and perhaps even willful self-destructiveness.

Since paternalistic laws force us, in ways that are more or less intrusive, to take care of ourselves in ways we may not wish to, they are countenanced by a liberal society such as ours

---

Adapted from the author's presentation at the New York Regional Conference on Medical Ethics at the Mount Sinai Medical Center on March 3, 1989. From the Department of Philosophy, Borough of Manhattan Community College and the Graduate School, CUNY, New York. Address reprint requests to the author, The Graduate School and University Center, 33 West 42nd Street, New York, NY 10036-8099.

(with its characteristic conception of the right of an individual to determine her own pursuits in her own way) only under certain conditions. The easiest paternalistic laws to accommodate are those which are seen as providing protections necessary for those who cannot protect themselves—children, say, or mentally retarded or incapacitated adults. Paternalistic restrictions in these cases—sometimes called cases of "soft" or "weak" paternalism—can be countenanced by a liberal society because they do not deprive the individual of any autonomous choice that he or she is in fact capable of making. The same might be argued when it comes to restricting the choices of those who, though not generally and permanently incapacitated, may be said to be acting in a particular situation under conditions that so compromise decision-making capacities that actions done under those conditions cannot be regarded as voluntary in that situation. Such might be the case, for example, with persons acting under conditions of coercion, intoxication, emotional duress, truly desperate economic conditions, and the like. Here the problem of whether paternalistic interference is morally permissible (or even required) involves questions that are both theoretical and practical: we need to clarify just how voluntariness is adversely affected, and we need to ascertain—which is not always easy—whether in a particular situation a particular person is acting under these adverse conditions (1).

But where there is no question that an adult is acting freely and voluntarily, then our paternalism, notwithstanding the laudable benevolence that motivates it, does violate that person's right to be left alone to freely pursue her life according to her own lights. To say this, however, is not to say that paternalism is necessarily unjustified in these cases. It may be, and it has sometimes been, argued that our duties of benevolence may sometimes override our obligation to honor a person's right to self-determination.

How, then, shall we regard legal paternalism with respect to drug abusers: is the law justified, for the good of the individuals concerned, in interfering with their lives for the purpose of preventing their harming themselves through drug abuse? There are several questions here:

1. Should we view the taking of drugs, at least in the absence of dependence on them, as sufficiently voluntary that legal interference with their use is morally problematic?
2. Is drug use—even that use which is sufficiently extensive that we might refer to it as drug "abuse"—sufficiently harmful to the users that even if the drugs are sought out and taken voluntarily, the state has the right (and perhaps even the duty) to paternalistically interfere with the lives of users?
3. Are our paternalistic drug laws effective with respect to their intended beneficiaries?
4. Is the particular way in which our present laws interfere with drug abusers for the sake of their own good—by criminalizing the use and possession of certain drugs—the least restrictive, and so—to a liberal society such as ours—the most acceptable among the possible reasonable alternatives?
5. Are the legal and social costs of our attempting, through our paternalistic laws, to effect good and prevent harm to drug abusers greater than the good they cause and harm they do?

These questions must form the basis for any discussion concerning the moral propriety of the legal measures adopted. (I am, of course, leaving out a great deal of what is important in legislation concerning drugs, chiefly the harms imposed by drug use on other individuals and on society as a whole. But these matters, important as they are, are not germane to the questions of paternalistic laws, laws whose concern it is to protect or further the interests of the drug user herself.) Considering the complexity, and the controversial nature, of the many issues embedded in these questions, it is not likely that we will easily be able to fashion comprehensive and respectable answers to these questions.

But I think that some of the things we will have to attend to in order to come to some answers are these:

First, we will have to distinguish between drugs: not all drugs—not even all illicit drugs—are harmful, even with heavy use; and not all dangerous drugs have the same deleterious effect on all users. The pharmacologic action of most drugs, so I understand, is a function not only of the properties of the drugs themselves, but of the users and the conditions under which the drug is taken. This fact is important because the justification of paternalistic interference will depend, at least in part, on the nature and magnitude of the harm it is designed to avert and the probability of the harm occurring. Since the extent and likeliness of the harm will differ with different drugs, and with different users of the same drug, questions concerning the permissibility of paternalistic legislation with respect to one drug may be answered quite differently from questions concerning the permissibility of paternalistic legal action with respect to other drugs.

Second, even if we were to establish, with respect to a given class of users, the harmfulness (or potential for harm) of a given drug or class of drugs, the moral propriety of legally interfering with the use of these drugs depends—if our motives are in truth paternalistic—on our ability to ensure that the consequences of our paternalism are, at the least, no worse for the drug user than use of the drug itself. Our paternalistic interference would have to do the user at least as much good as the relief, however temporary, provided by the drug itself. This is a tall and, perhaps, an impossible order, considering the often desperate circumstances that make a life of drug use more bearable than the life from which it provides escape. (Of relevance here are the reports in the psychiatric literature on the "preferential use of drugs" [2], that is, the selection, by habitual users, of specific drugs to provide what may be regarded as self-medication for a range of psychiatric problems and painful emotional states [3].)

Third, paternalistic impositions concerning drugs (as, indeed, paternalistic impositions for any reason) are justified only if there is a reasonable chance that they will be effective in achieving their intended benefits. However, our present laws—which criminalize the possession, sale, purchase, and use of various drugs—have no record of success. They have not helped drug users and abusers, first and foremost, because they have not deterred them from using drugs. This is, some say, just what we should have expected: drug users who are not deterred by the dangers of drug use itself are not likely to be deterred by the risks of criminal involvement that our legislators have made a necessary corollary of drug use (4). (Even Operation Pressure Point—a program regarded as highly successful in ridding New York City's Lower East Side of blatant and pervasive drug trafficking—probably succeeded not in reducing the number of the city's drug users, in absolute terms, but only in shifting the drug market from one location to another [5].)

But aside from the fact that legal paternalism will not find justification if it does no good, it surely will not be justifiable if it actually does harm. Sadly, this is not a moot issue with respect to our current laws, for in criminalizing the purchase and use of certain drugs, they have made those who might otherwise have been drug-using, but nonetheless law-abiding citizens into de facto criminals and have created the circumstances that have led to a black market in the forbidden drugs that prices them beyond the reach of what most legal sources of income can provide. These results, in turn, have encouraged the drug user, especially if he is drug dependent, into extralegal activities to support his needs.

The harms have been colossal: Criminalizing drug users has created a social underclass for whom being outside the law is an inevitable fact. Apprehending and processing drug users through the channels of our criminal justice system (arrests, trials) has overburdened it. The overburdened system has failed to bear the weight of its ever-increasing workload; the resulting cynicism toward agencies of law enforcement—in both drug users and the general public—has been socially costly. Police have become subject to the corruption of big money, and organized crime has been given an opportunity for exploitation.

Yet all of this is not to say that paternalistic legislation is out of place with respect to drugs. If we can keep our focus on the fact that, as one writer has put it, "drugs are not the problem, people turn to drugs because they have problems" (6), we will remember the connection between being paternalistic and being humanitarian. This may help us to formulate laws that provide genuine assistance to drug users—that promote their welfare rather than bring them to ruin.

**Laws and Moral Ideals.** The heavy use of many drugs (at least for the majority of users) is harmful, indeed quite harmful, but nonetheless that use is considered none of the law's business. We drink as much coffee and swallow as many aspirins as we desire, and so long as we do not drive under the influence of alcohol, there is no legal limit to the extent to which we may put ourselves under its influence (at least in private). We regard it as beyond the law's rightful business to legislate (though it may comment on) the propriety of our smoking ourselves into a state of emphysema, and the suggestion that common table salt should be offered on the market only under restrictive paternalistic legislation appears almost funny, even though there is a clear connection between the incidence of hypertension and the amount of salt we use.

The excessive use of many drugs is acknowledged as harmful but is, nonetheless, regarded as outside the legitimate domain of paternalistic legislation. This should alert us to the fact that benevolence alone does not explain our willingness to legally restrict our fellow citizens' behavior with respect to drugs. Perhaps, behind the guise of paternalistic interest, something else entirely underlies the criminalization of the use of some drugs.

It may be that we associate certain types of drugs with certain socially, economically, or racially defined groups of people. It may be that we associate certain types of drugs with certain types of effects that our society—with its subscription to the puritan ethic (equating industry with merit)—regards as hallmarks of a morally dissolute life. Use of illegal drugs effects drug-induced calm, diminished work performance, diminished incentive for industry and participation in a "productive life," a reduction in initiative to "better" ourselves. Contrast these effects of the drugs we prohibit with the effects of those drugs—even the especially addictive ones like nicotine and caffeine—for which we have great social and legal tolerance. Use of legally tolerated drugs sharpens rather than dulls our awareness, heightens rather than depletes our energy, and increases rather than lessens our productivity.

This contrast suggests that we are using the law not to help us take better care of ourselves, but to underscore our view of moral merit. What we have here, then, is not merely the use of law to protect or promote the interests of the subjects of our benevolent interventions, but the use of the law to legislate a moral ideal, to which the subjects of the law may or may not subscribe.

## References

1. H. L. A. Hart has pointed out, in his discussion of John Stuart Mill's view on justified paternalistic interference, that Mill's conception of human nature is wide of the mark. We are, Hart believes, more susceptible to subtle and wide-ranging factors that diminish our capacity for full voluntary action than Mill allowed. Hart calls our attention to such factors as inner compulsions, social pressures, failure to appreciate consequences, being in predicaments likely to cloud judgment, etc. See Hart HLA. Law liberty and morality. New York: Oxford, 1963:32–33.
2. Milkman H, Frosch WA. On the preferential use of amphetamine and heroin. J Nerv Ment Dis 1973; 156:242–248. See also Wieder H, Kaplan EH. Drug use in adolescents: psychodynamic means and pharmacogenic effect. Psychoanal Study Children 1969; 24:399–431.
3. Edward J. Khantzian notes that "addicts discover that the short-term effects of their drugs of choice help them to cope with distressful subjective states and an external reality otherwise experienced as unmanageable or overwhelming." Khantzian EJ. The self-medication hypothesis of addictive disorders: focus on heroin and cocaine dependence. Am J Psychiatry 1985; 142(11):12–63.
4. Kaplan J. The role of the law in drug control. Duke Law J 1971:103.
5. Zimmer L. Operation pressure point: the disruption of street-level drug trade on New York's Lower East Side. New York: The Center for Research in Crime and Justice at New York University School of Law, 1987. Also relevant is the report on the Rockerfeller Drug Laws issued by the Department of Justice. It shows that the increased penalties instituted by these laws reduced neither the number of drug dealers nor the supply of drugs available in the state (U.S. Dept. of Justice, 1978).
6. Whitlock EA. Drugs, morality and the law. Brisbane: University of Queensland Press, 1975:50–81.

# Pharmacology of the Drugs of Abuse and the Development of Public Policy

Robert B. Millman, M.D.

The recently renewed debate on alternatives to the current drug prohibition policies in the United States is valuable. The social and economic costs of drug prohibition policies have been enormous, and we must continue to seek more effective responses to the drug problem. At the same time, the well-meaning arguments for legalization offered by social scientists, historians, ethicists, and politicians have been incomplete (1, 2). Many people, particularly nonclinicians, tend to think of drug abuse and dependence generically, to abstract the drug abuse behaviors, and consider the impact of the "drug problem" primarily in economic or social terms. The relationship of the pharmacology of specific drug classes to abuse patterns and to the adverse sequelae of these behaviors tends to be ignored. Issues such as the psychoactive effects of the various drug classes, tolerance, physical dependence, reinforcement, acute and protracted withdrawal syndromes, and the relationship of psychopathology to individual drug-taking patterns and the development of adverse reactions are generally not considered in any detail.

Each drug class and each drug is different; each has its own ethos, pattern of abuse, dependence liability, and adverse reactions. Pharmacologically, some drugs are much more dangerous than others for a variety of reasons, and increased availability and use pose significant danger to the individual and society (3). For example, use of cocaine, particularly in the "freebase" or "crack" preparation, is more likely to be associated with compulsive use and dangerous sequelae than marijuana or even the hallucinogens. The pharmacology of a drug is not the sole determinant of dependence liability or adverse sequelae; despite the pharmacologic dependence liability of a particular drug, most experimenters or casual users do not get into trouble with any of the drugs, are able to control use, and do not become dependent.

Zinberg (4) has noted that individual drug-taking patterns are influenced by set and setting in addition to pharmacology. That is, personality characteristics, psychopathology, and expectations, as well as the environment and social setting in which the drug is taken, influence the psychoactive effects, adverse reactions, and likelihood of dependency. He further suggested that social sanctions and rituals are developed in most societies to help people control drug use. For example, in the mountains of Colombia and Peru, coca leaf chewing was integrated into the social fabric of certain groups, and adverse sequelae were minimized. Similarly, among the hill tribes of Thailand, opium smoking came to be associated with power and wisdom and not outcast status (4).

It has been suggested that if we legalize drugs and teach people how to use them properly, the negative impact of both prohibition policies and the drugs themselves might be minimized. There are problems with this argument. In earlier or less affluent societies, the only drugs available were indigenous to the region, and the culture could learn to control them. Currently in our

---

Adapted from the author's presentation at the New York Regional Conference on Medical Ethics at the Mount Sinai Medical Center on March 3, 1989. From the Department of Public Health, Cornell University Medical College, New York. Address reprint requests to the author, Saul P. Steinberg Distinguished Professor of Psychiatry and Public Health, Cornell University Medical College, 411 East 69th Street, Room 305, New York, NY 10021.

society it is not unusual for a clinician to evaluate a patient who uses cocaine from Peru, heroin from Cambodia, cannabis from Jamaica or Kentucky, diazepam from Michigan, and vodka from Russia. People in Western society have access to a bewildering array of psychoactive substances and it is difficult to imagine responsible use patterns developing for all of these in the foreseeable future. Few rituals or sanctions associated with, for example, the use of heroin or cocaine would help people integrate drug-taking patterns into a reasonably productive mode of living.

In addition, given a social situation in which large numbers of people are poor and culturally deprived, as in many inner-city areas, coupled with the increased availability of drugs at lower prices, what is likely is a considerable increase in drug use and adverse sequelae, both individual and social. A significant portion of the poor population does not perceive drug use as particularly dangerous or likely to interfere with social and vocational status and aspirations. In fact, the drug use of many people, particularly the youthful and the economically or culturally impoverished, can be considered purposive, goal-directed activity. These people feel powerless, perceive themselves as having little to lose, and believe that drug use is a reliable, predictable way to gain power, control, and pleasure in the absence of other reasonable alternatives; it is not surprising that dangerous drug use patterns have become endemic in certain areas. Then too, most people believe they can control their drug use, particularly early on, and that they will not get into trouble (5).

The situation with respect to alcohol is instructive in this regard. Alcohol is clearly one of the most dangerous psychoactive drugs in terms of dependence liability and the development of untoward side effects. Since it is a legal drug, use patterns are importantly influenced by social sanctions, such as "Don't drink alone" or "Don't make a fool of yourself," and drinking is limited by social rituals such as the cocktail party. Most people are able to control their alcohol use; they are social drinkers. The enormous number of people who drink, however, reflects the psychosocial and metabolic risk factors of alcohol. There are approximately 10 million alcohol addicts, many suffering from alcohol-related problems, including cirrhosis of the liver (6). Automobile accidents, assaults, homicides, and suicides influenced by alcohol are a much more significant public health issue than are the consequences of all the illicit drugs combined.

## Pharmacology of the Illicit Drugs

**Opiates.** It may be useful to briefly consider the pharmacology of some of the most frequently abused drugs with a view to assessing the impact of legalization. The opiates or narcotic analgesics include heroin, morphine, methadone, and codeine. Psychoactive effects include a sense of well-being, calm, and sedation. During the period of intoxication, users experience decreased appetite for food and sex and tend to be passive and calm. Violent behavior is unlikely when users are "high". Physically, opiates are relatively nontoxic unless the dose is excessive or use involves unsterile conditions. Productivity is generally decreased, although some people are able to function while intoxicated.

Opiate use is, however, associated with a high degree of tolerance—the need for increasing doses to regain the initial effects—and the development of physical dependence, defined as an altered physiologic state induced by repeated administration of a substance and requiring continued administration to prevent the characteristic withdrawal syndrome (3). The continual need to increase the dose of heroin, for example, to get desired effects leads to expenditure of more money for more drug and the tendency to maximize effects by progressing from sniffing the drug, to intramuscular use, to intravenous use. The opiate withdrawal syndrome, while not life threatening, is experienced as profoundly dysphoric. It is marked by depression or anxiety, abdominal cramps, diarrhea, running nose, and goose bumps. This syndrome occurs three to four hours after the last dose of heroin, and addicts will do just about anything necessary to get another dose of the drug. If the drug is not available, suicidal feelings are not unusual; addicts feel as if they cannot go on. Heroin-dependent people are likely to commit any sort of crime to get money for the drug; they may be violent incident to these efforts, though violent crime is not a preferred mode of behavior. Acquisition and use of increasing amounts of the drug becomes the primary life concern for the addict with an opiate dependency (5).

Great Britain attempted to provide heroin by prescription until the mid 1960s, when it became clear that the program was not working. During that period a youthful drug culture emerged with markedly increasing demands for drugs. Because of the short duration of action of the drug and the progression of tolerance and physical dependence, these English addicts tried to get as much drug as they could from their private doctors and, subsequently, from government clinics. At the same time the programs attempted to keep doses as low as possible, so that many addicts secured their desired supplies illicitly. In recent years, the number of heroin addicts in Great Britain has significantly increased, largely due to increased supply of the drug, first from Iran and then from Southeast Asia. Currently, most maintenance programs in England provide methadone rather than heroin, recognizing that patients can be stabilized at a daily oral dose that prevents the development of withdrawal symptoms, alleviates opiate craving, and blocks the euphoric effects of heroin. If the British were unable to maintain a few hundred or a few thousand addicts on heroin, it is unlikely that American government or industry could accomplish this. Providing heroin to addicts in governmental programs or drug stores is likely to lead to more illicit use, not less (7).

The situation becomes even more problematic when the natural history of heroin dependence is appreciated. After cessation of opiate use, whether with or without a treatment program, a large percentage of addicts relapse. The traditional view was that addicts had a preexisting personality disorder or other psychopathology that they self-medicated with heroin, and the persistence of the condition led to relapse after a period of abstinence.

Dole in 1965 (8) suggested a different mechanism for relapse: that long-term opiate use creates a metabolic change, a neurochemical deficit that persists for long periods after cessation of drug use and may serve as a powerful reinforcement to resume drug taking. More recent basic scientific research has begun to lend credence to this important idea (8). The discovery of specific receptors for opioid substances and the isolation of the agonists themselves, the endorphins and related neurotransmitters, suggests the possibility of a neurohumoral feedback system, or a persistent receptor disorder, as etiologic in both acute and protracted withdrawal symptoms. Under resting conditions when exogenous opiates such as heroin are administered, the production or secretion of the endogenous neurotransmitter may be suppressed. Organisms attempt to maintain homeostasis, in this case a stable level of circulating neurotransmitters. Continued drug taking may lead to continued suppression of this system, which may be involved in pain control and mood maintenance. Perhaps the cessation of exogenous opiate use, by down regulation of the modulating receptor agonist system and suppression of endogenous opioids, contributes to narcotic tolerance and acute and long-term abstinence symptoms (8).

Long-abstinent addicts experience variable depression, apathy, loss of energy, increased anxiety, insomnia and powerful persistent cravings that may be related to "protracted abstinence syndrome." These metabolic changes may be responsible for the tendency to relapse.

The increased availability of high-grade heroin through legalization might increase the number of heroin users and exacerbate the pharmacological issues of significant tolerance and physical dependence. Though drug-related crimes might decrease initially, a generation of people at risk might find their way to heroin dependency and profound neurochemical change. Women who use heroin may also pass drug-related problems on to the next generation. Pregnancies associated with heroin use often produce low-birth-weight children who are physically dependent. Whereas the acute withdrawal syndrome in these infants is easily treated, the likelihood exists that these infants may have altered neurotransmitter balance; some may be found to have chronic cognitive or emotional deficits and impaired ability to lead rewarding and productive lives.

**Cocaine.** Cocaine and the amphetamines are psychomotor stimulants. Effects include increased wakefulness and alertness and decreased sense of fatigue. Mood is elevated, accompanied by increased confidence, energy, and initiative. Performance is improved acutely, particularly when it has been impaired by fatigue.

When "snorted," until recently the primary mode of administration in this country, effects are perceived within minutes and persist for approximately 30–60 min. When cocaine is taken intravenously, users experience an intense and immediate "rush" that passes within moments. Most recently, an efficient method of snorting cocaine "freebase" has become popular. Cocaine hydrochloride decomposes if heated to the point of vaporization; conversion to the alkaloid form lowers the vaporization temperature by about 100°F and allows it to be smoked efficiently. "Crack" is a widely available form of cocaine base. Psychoactive effects of smoking are virtually indistinguishable from the intravenous mode and in both routes of administration extremely high blood levels occur rapidly. Some degree of tolerance to the effects of cocaine occurs, although the extent of tolerance is limited compared to the opiates. Physical dependence is thought to occur, though it is less important to the addictive process than that induced by the opiates.

If decriminalization is to be considered, the critical issues are the reinforcing and psychomimetic properties and adverse sequelae of the drug (9–12).

Cocaine is generally used in "runs" of several hours to several days or more. These episodes are marked by intense involvement with the drug and little interest in food, sleep, or other activities. Performance declines, and users tend to experience increased anxiety and paranoia as the run continues. Continued use of high doses of the drug, particularly by the intravenous or smoking route, leads to a well-defined psychotic picture indistinguishable from functional paranoid psychosis. At this stage users may be disorganized and engage in stereotyped repetitive rituals—compulsive masturbation, compulsive checking that a door is locked, that they are not being watched or listened to (13).

In rats, monkeys, and humans, cocaine is considered to be the most reinforcing drug (9). The more rapid the onset of effects and the higher the blood level obtained, the more powerful the reinforcement. Whereas compulsive use may be induced by any route of administration, smoking and intravenous use are more reinforcing than the others. The specific mechanisms of reinforcement are currently under study, though some insight into the phenomenon may be gained from clinical observation. After the initial rush, users experience a rapid decline in the desired psychoactive effects. In essence, with cocaine, on the way up, users feel good, though on the way down they experience dysphoria. With heroin, craving does not occur for three to four hours after the last administration of drug, when withdrawal is beginning. With cocaine, craving occurs almost immediately after use. Users will smoke crack or inject the drug every 10 to 20 minutes. When supplies run out, a user in the thrall of a cocaine binge will ardently attempt to replenish supplies (9, 11). In addition, high doses have been associated with severe adverse medical sequelae, including cardiovascular and cerebrovasular accidents and sudden death (10, 12).

It has been well documented that mothers will abandon or even abuse their children while under the influence of cocaine. Also increasingly well documented among cocaine-using pregnant women is significant fetal loss; in addition, the children born to cocaine-abusing mothers may suffer severe adverse sequelae (14, 15).

Thus the problem of legalizing or decriminalizing cocaine is that it is so powerfully reinforcing that at high-dose use, a significant proportion of users will lose control over their craving. Because people using cocaine become agitated and anxious and may be psychotic, increased availability of the drug in relatively pure form is likely to lead to a marked increase in violent and irrational behavior.

Proponents of legalization believe that the remarkable increase in drug-related crime in the inner city is primarily related to the buying and selling of the drug, that is to the economic situation caused by prohibition policies. Clinicians, however, recognize that the pharmacology of the drug, particularly its reinforcing properties and its psychotomimetic effects, are critical to understanding the impact of the drug on individuals and society. What would the societal response be to the agitated angry user in the throes of a run presenting herself or himself for more drug? How would we deal with teenage users? How would we respond to cocaine addicts who got extra doses outside the accepted system?

**Cannabis, Hallucinogens, and the Sedative Depressants.** Whereas use of cannabis, the hallucinogens, and the sedative depressants may be fraught with serious disability, arguments against legalization would depend less on pharmacologic grounds than on social and ethical concerns. It could be argued that these drugs, more than other drugs except for alcohol, endanger the user rather than the society. Although violent or criminal behaviors precipitated by cannabis or the hallucinogens occur, they are rare. Perhaps drugs that induce lassitude or passivity are less problematic for society, although they may be destructive for the user (5).

## Summary and Conclusions

It is fascinating that the morality, ethics, and effectiveness of current drug prohibition policies are being criticized in an era when drug use is declining in most segments of society. The current era is also marked by increasingly restrictive policies against drug use and users, including much enhanced interdiction efforts, calls for tougher penalties and more jails, and extensive drug screening programs at the work site. Summary dismissal of workers or students who use drugs is increasingly advocated.

Two distinct and opposing currents advocating policy change have emerged, one for legalization, the other for greater restriction. The current inner-city rise in drug use, particularly of cocaine and "crack," and the remarkable increase in drug-related crime is probably providing the impetus for both lines of thinking. Yet both views appear extreme and neither seems to offer a promising social policy.

It is critical that we continue to attempt to develop effective strategies for reducing the impact of drug taking on the individual and society. Whereas I am sensitive to the ethical and legal perspectives of the proponents of legalization, pharmacologic and social issues persuade me that legalization of heroin and cocaine would lead to a marked increase in drug dependence and an increase in drug-related disability and crime. In response to the idea that the legalization and taxation of drug use would provide resources for more effective research, prevention, and treatment, my sense is that we should not wait for legalization but should embark on that long and expensive course immediately. More expensive and restrictive interdiction and criminal justice efforts have proven to be wasteful, not particularly effective, and damaging to individuals and society.

Rather than legalizing drugs, we should continue to attempt to reduce demand with effective prevention efforts, including the provision of realistic, rewarding alternatives to those at risk, the youthful and the psychosocially disadvantaged.

Drug use should be medicalized though not encouraged, and more effective and attractive treatments should be developed for all who need them. We might construct a hypothetical continuum of drugs and drug classes ranging from the less dangerous to the more dangerous and use it as a guide for decisions about decriminalization. Using such a guide, alcohol and the sedative depressants might continue to be legal. Marijuana and hallucinogens are less apt to cause dependency and adverse reactions and so should be associated with lesser legal sanctions. Decriminalization of heroin would not be feasible, although drug-related crime might decrease as a result of such a policy change. Cocaine and "crack" and amphetamine and "ice" cannot be legalized because of the nature of their reinforcement, associated dependency, and toxic properties. Certainly discussions of public policy with respect to drug use should be more sensitive to these pharmacologic issues.

### References

1. Nadelson EA. Drug prohibition in the United States: costs, consequences and alternatives. Science 1989; 245 (Sept. 1)
2. Kerr P. The unspeakable is debated: should drugs be legalized? The New York Times, p. 1, May 5, 1988.
3. Millman RB. Drug abuse and dependence. In: Francis AJ, Hales RE, eds. Psychiatry update, American Psychiatric Association annual review, vol. 5. Washington, DC: American Psychiatric Press, 1986, 120–227.
4. Zinberg NE. Drug set and setting, the basis for controlled intoxicant use. New Haven, CT. Yale Univ Press, 1984.
5. Millman RB, Botvin GJ. Substance use, abuse and dependence. In: Levine MD, Carey WB, Crocker AC, Gross RT, eds. Developmental behavioral pediatrics. Philadelphia: WB Saunders, 1983, 683–708.
6. Maryland U.S. Department of Health and Human Services. Sixth Special Report to the U.S. Congress on Alcohol and Health, Jan. 1987.
7. Wilson JQ. Against the legalization of drugs. Commentary 1990; (Feb. 26):21–28.
8. Dole VP. Implications of methadone maintenance for theories of narcotic addiction. JAMA 1988; 260(20): 3025.
9. Johanson C, Fishman MW. The pharmacology of cocaine related to its abuse. Pharmacol Rev 1989; 41(1):3–45.
10. Sbriglio R, Millman RB. Emergency treatment of acute cocaine reactions. In: Washton AM, Gold MS, eds. Cocaine abuse: recent trends and clinical perspectives. New York: Guildford, 1987, 87–95.
11. Miller NS, Gold MS, Millman RB. Cocaine. Am Fam Physician 1989; 39(2):115–120.
12. Tardiff K, Wu J, Gross E, Millman RB. Analysis of cocaine-positive fatalities. J Forensic Sci 1989; 34(1):53–63.
13. Millman RB. Evaluation and clinical management of cocaine abusers. J Clin Psychiatr 1988; 49:2(Suppl)27–33.
14. Chasnoff Ira J, et al. Temporal patterns of cocaine use in pregnancy: perinatal outcome. JAMA 1989; 261 (12):1741–1744.
15. Doberczak TM. Neonatal neurologic and electroencephalographic effects of intrauterine cocaine exposure. J Pediatr 1988; 113(20):354–358.

# Drug Policy in New York City

Stephen C. Joseph, M.D. M.P.H.

The most pressing public health issue facing New York City is the epidemic of drug abuse, which is inextricably interconnected with many of the city's most intractable social problems.

**AIDS and Drugs.** Fifty years from now the AIDS epidemic may appear as but one expression of the larger epidemic of substance abuse in our city and our country. The gay community has reduced the spread of new infection, but among IV drug users the virus continues to spread virtually unabated. In the first three months of 1988, new cases of AIDS among IV drug users in New York City surpassed the number of new cases among gay men. In 1989 up to 8% per year of IV drug users seroconverted whereas the rate of seroconversion among men who have sex with men is 1% per year.

This epidemiologic shift has a number of implications for New York City. The increasing incidence of AIDS and HIV-related sickness and death in IV drug users and their sex partners and children is having an overwhelming impact on this city's poor and minority residents. Blacks and Hispanics, disproportionately represented among the poor and therefore also among IV drug users, already account for close to 60% of the reported 19,000 cases of AIDS: 85% of women, and over 90% of children with the disease come from these two minority groups. The AIDS epidemic is tracked straight at minority and poverty neighborhoods (and, of course, at the hospitals that serve these communities).

The spread of HIV infection among needle-users is not confined to those who shoot heroin. Cocaine injection has joined IV heroin use as a major route of HIV spread. Those who use cocaine intravenously inject more frequently than heroin users, and are at great risk of infection through shared needles. Additionally, crack and the hypersexuality associated with it, including sex-for-drugs transactions, has led to increases in HIV transmission and heterosexual HIV infection, as well as increases in genital ulcer disease. Clearly, the future of AIDS in New York City depends on our ability to slow the spread of HIV from, and among, substance abusers.

**Children and Drugs.** Beyond the AIDS epidemic, substance abuse is wreaking havoc on New York's children and families in many other ways. In 1986, half of the infants registered in the Department of Health Infant Health Assessment Program (IHAP) were identified by maternal substance abuse either by itself or in conjunction with other health risk criteria. In some Health and Hospital Corporation hospitals, this risk factor accounted for up to 70% of the infants registered: These infants represent over 7% of the total live births in the 12 IHAP hospitals.

Maternal substance use, cocaine use in particular, can contribute to premature deliveries and low birthweight. Cocaine use by pregnant women increases the risk of life-long harm to unborn children, and puts the newborn at a greater risk of dying during the first year. The mortality rate for infants born to substance-using women, predominantly mothers who abuse cocaine, is as high as 2.5 times that for all infants in the city. Reports on New York City births between 1978 and 1986 show that maternal cocaine use has increased 20-fold. In 1986, cocaine was mentioned in two-thirds of the reports of New York City births where the mother was identified as a substance abuser. More recent figures are expected to show a continuing increase in cocaine use by women during pregnancy.

The list of public health ills connected with substance abuse continues. Drug abuse, particu-

---

Adapted from the author's presentation at the New York Regional Conference on Medical Ethics at the Mount Sinai Medical Center on March 3, 1989. At the time of this discussion the author was Commissioner, Department of Health, New York City. Address reprint requests to the author, Dean, University of Minnesota School of Public Health, Box 197 Mayo, 420 Delaware St. SE, Minneapolis, MN 55455.

larly among men, is in part responsible for the rise of tuberculosis in New York City. And crack is contributing to the growing incidence of sexually transmitted diseases, as well as the spread of HIV.

**Treatment and Prevention.** Substance abuse is not a medical problem, but a social problem with medical implications. Treatment and prevention of substance abuse are critical, but we can and must do more to stop this epidemic. We will not make a dent in the enormous public health problems that face us unless and until we mount a multidisciplinary, all-out war on drugs. It must include interdiction at the international level; law enforcement at all levels; more education programs; increased, liberalized, on-demand methadone maintenance and massive detoxification programs for those already addicted; and availability of clean needle exchange, as called for by the National Academy of Science. The step I believe we should not take is legalization of drugs. In an urban area such as New York City, with a large population of vulnerable young people—many already reeling from the effects of poverty and disease associated with the "new morbidity"—legalization would boost our current substance abuse crisis to another, almost unthinkable level of magnitude.

Although no single strategy will by itself be fully effective against the complex problem of substance abuse, greatly increased availability of drug treatment programs must be our first line of defense. Existing drug treatment programs in New York City can accommodate only 35,000 of the city's 200,000 IV drug users. The State Division of Substance Abuse Services estimates that twice the number, 70,000, would seek drug treatment if it were available on demand. Despite the public health emergency, there has been no expansion of drug treatment facilities in New York City within the past ten years. Despite the magnitude of the AIDS crisis, only a slight increase has occurred in the number of treatment slots available—regrettably the state has not responded to our proposal for using several Health Department sexually transmitted disease clinics as possible sites for new evening drug treatment programs. We must increase programs for treatment on demand, including liberalization of methadone maintenance programs, inpatient detoxification programs, and inpatient and outpatient residential facilities.

The situation is even more complicated by our total lack of proven effective mass-applicable treatment modalities for cocaine addiction. Much current cocaine treatment is patterned after drug-free treatment for heroin abuse without adjustments for the many differences between heroin and cocaine. Further, no clear evidence suggests that any one cocaine treatment is superior to another, or that specific patients should be matched to particular psychotherapies.

At present, virtually our only options for treating cocaine and crack addiction are intense individual and group psychotherapy and residential care. The success of these methods largely depends on a supportive network of drug-free families, friends, and environments to which the patient can turn when the craving for relapse is strongest. Creating such settings is costly and not enough are available to meet the demand. Too often addicts are left with disorganized or nonexistent support systems.

Although evidence is mounting that pharmacotherapy may be effective in the treatment of cocaine abuse, the critical need is for national biomedical research to develop a chemical agent to block the effects of cocaine, much as methadone substitutes for heroin. The need for such a cocaine blocker for use with psychotherapy and prevention programs is particularly urgent in urban areas like ours, where health, law enforcement, and social services have been overwhelmed by the crack epidemic.

Faced with the realities of uncertain treatment for cocaine addiction and a general dearth of treatment places, the Department of Health has developed many programs and services, especially those related to AIDS, aimed at prevention of drug abuse. These efforts include an expanding network of voluntary, anonymous risk-reduction counseling and counseling-based HIV antibody testing sites, and more and better street-worker outreach in communities of heavy drug use to educate both IV drug users and their sex partners on substance abuse issues.

Our earlier efforts to bring prevention information to the public directed resources into frank television, print, radio, and subway advertising campaigns. These efforts also brought the AIDS prevention message to drug abusers and their partners, to nonabusers, heterosexuals, and adolescents. These efforts continue today, but the major focus has now shifted to funding 30 community-based organizations which provide substance abuse counseling, education, and information. These community-based organizations have credibility, cultural familiarity, and direct accessibility to the target populations. They present unparalleled opportunities for street-level education and advocacy.

We still have miles to go in our efforts against the tremendous problems and dilemmas of substance abuse. But addressing the problems of substance abuse also provides opportunities for resolving other interconnected social problems.

# Drug Use and Public Policy

JOHN BRIAN MURTAUGH

I SPEAK here as an angry man. I am angry about a lot of things so I do not share Dr. Stephen Joseph's measured voice of reason. I am the New York State Assemblyman representing northern Manhattan: Washington Heights, Inwood, Marble Hill, and the George Washington Bridge communities. Part of my district, Washington Heights, is one of the crack capitals of New York. I have been working on the front line of the drug problem for a long time, first as a community activist and now as an Assemblyman. Many people come into my office looking for treatment, and my office staff members spend as much as half a day trying to find treatment spots for them. I created the New York State Assembly Subcommittee on Alcoholism and for a number of years I served as its chairman.

Although I have acquired some knowledge of alcoholism and substance abuse through my work, I believe that my greatest asset as the new Chairman of the Assembly Committee on Alcoholism and Drug Abuse is that I realize what I do not know. Many physicians lack precisely this knowledge. Most see addicts at one time or another for medical care, but I doubt that most doctors recognize addiction in patients. Even after an examination, I am not certain that a doctor would know what to do if the problem was correctly identified. I do not believe that the individual physician is entirely to blame; I think this failure to recognize and address addiction is the fault of the entire medical training system, from medical schools through residency training.

**Drug Abuse and Alcoholism.** One thing I do know is that when we talk about drug abuse we must also talk about alcohol. Almost two-thirds of the addicted people in the United States are addicted to a combination of alcohol and prescription or non-prescription drugs, particularly cocaine. Two years ago a survey done of about 27,000 students from 9th through 12th grades selected at random to represent 1.5 million students throughout the State of New York, showed that about 21% did not touch alcohol or drugs, about 9% had tried only drugs, 19% had tried only alcohol, and 53% had tried drugs (primarily marijuana) and alcohol. Some of these students were using both alcohol and drugs at the same time. This last group appears to be most at risk for addiction. Whatever public policy we create, we will have to address these young people and their problems.

I am angry right now about what I see developing in Albany. The Governor of the State of New York declared a war on drugs, and then immediately cut the alcoholism budget by over 15% for 1989. [In 1990 a substantial budget increase was voted but it was never implemented. The 1989 status quo continues.] It is ironic, because alcoholism and drug abuse go together; you cannot talk about one unless you talk about both of them. At present we have an inadequately supported multitiered approach to treatment: methadone maintenance (which works in some circumstances), therapeutic communities (which work in many circumstances), and a variety of out-patient treatments. But to be successful we must expand those programs and get more serious about relapse treatment. If we start from the premise that treatment can work, we must find treatment modalities that work for different populations and types of addiction.

Also, I am upset about the way physicians typically take a one-dimensional view of addic-

---

Adapted from the author's presentation at the New York Regional Conference on Medical Ethics at the Mount Sinai Medical Center on March 3, 1989. From the New York State Assembly Committee on Alcoholism and Drug Abuse, Albany. Address reprint requests to the author, Chairman, New York State Assembly Committee on Alcoholism and Drug Abuse, Room 746 LOB, Albany, NY 12248.

tion. Doctors often assume that addiction is a single disease for which the patient is to blame. And they assume that if they get the patient into treatment and the patient then falls off the wagon, it is undoubtedly the addict's fault and, therefore, relapse treatment for the addict is neither useful nor deserved. Such thinking is simply wrong. Often, relapse treatment is cheaper and more effective because it does not have to start from ground zero. We are now starting to talk about mandating insurance coverage for in-patient treatment for alcoholism. One of the things I am including in this legislation is a program to encourage care providers to do relapse treatment. If this legislation can be implemented for alcoholism it will, we hope, at some point also be available for substance abuse. If someone relapses, treatment should be offered again by the same provider whenever possible.

As a society we have been dealing with our drug problem predominantly through a law-and-order approach. One encouraging sign of the times is that even the people in law enforcement admit this approach is not working. Recently, on a WCBS half-hour forum, I heard a top New York City law enforcement official explain that dealing with drugs must be like a stool with three legs: one leg is interdiction, law and order; the second is prevention; the third is treatment. I was encouraged to hear that comprehensive understanding from a police official. This view needs to be repeated loudly and more often so we can get that message across to the public. When it comes to drugs, most people unfortunately still say that you have to deal with the disease as a crime.

**Budgets and Treatment.** I have the following goals as the chairman of the Assembly Committee on Alcoholism and Drug Abuse. My first is very short-range. The 1989 state budget has an $8.5 million cut in the alcoholism budget, which translates to a decrease of about $22 million because of the corresponding loss of funds from local communities and insurance companies. My first priority is to reinstate these funds.

My second goal is for the State of New York to adopt a policy of drug treatment on demand. It is absurd to declare war on drugs and then have people coming in ready for treatment, wanting to go into treatment, only to be placed on a long waiting list. Addicted people do not do well on waiting lists. When a treatment slot finally becomes available, the addict is often no longer around. So when the "magic moment" is there, treatment slots must be available as a matter of public policy. It is going to cost some money to provide this treatment, but not providing it will cost a lot more. In-patient treatment costs about $18,000 a year, not including the construction cost of new beds. But it costs society $33,000 a year in crime to have an addict out on the street. Right now, the City of New York is jailing about 9,000 people per week and holding them one to five days. Many of these people are substance abusers and for them this contact could be turned into the moment of intervention and an opportunity to begin treatment.

Many addicts want treatment. Joseph has noted that among the limited number of people in the City's short-lived needle exchange program, many talked about getting treatment. Significant numbers of addicted people with AIDS seek treatment. It is as if learning their diagnosis ignites that spark of life within them. We have to be prepared to take advantage of these magic moments and provide opportunities for treatment.

Treatment is cheaper than all other alternatives. Today it presumably costs anywhere from $25,000 to $35,000 a year to maintain a prisoner in the state prison system. (I believe the true cost is higher.) It costs $110,000 to build each cell. Compare that to $18,000 for drug treatment per year, and beds that cost $35,000 to $45,000 to build. It is cheaper to treat anyone who seriously seeks treatment. Instead the public wants revenge through tougher laws and minimum sentences. Even though people are reluctant to build more jails, we will have to. The federal courts will not let us get away with overcrowding; and the politicians are trying to get away with talk.

Right now the City of New York and the state are arguing over who is responsible for funding treatment. Treatment, however, is a three-way responsibility belonging to the federal government, state government, and local government. The City of New York is treating people presently within the hospitals and within the jail system. That service could be expanded enormously and supplemented with a component to serve people when they leave the jail. Perhaps a series of halfway houses for these newly recovered addicts would be useful, especially since many people coming out of the jails would, otherwise, become homeless.

We have lost opportunities for treatment within the Health and Hospital Corporation system. I remember seeing an application for an alcoholism program, in a prestigious inner-city voluntary hospital, stating that 30%–50% of the patients from the inner-city neighborhood probably abused alcohol or else abused a combination of alcohol and other drugs. However, a study of its discharge plans showed that only 3% of plans mention treatment for alcoholism as part of the discharge plan. That is outrageous. For any other disease (congestive heart failure, tuberculosis, AIDS) the hospital would have lost its operating certificate if it so often failed to diagnose and list treatment on the discharge plan. I am sure the same type of oversight occurs at most hospitals. We are missing opportunities.

**Related Issues.** Beyond my goals of restoring the state's alcoholism budget and providing for treatment on demand, I have several other goals as chairman of the Committee on Alcoholism and Drug Abuse. The State of New York must deal with what is happening to family structure. It must deal with the child abuse and child neglect associated with drug use. In ten years the badly abused 5-year-old will be a dehumanized 15-year-old capable of launching his own reign of terror. The time bomb is ticking, so it is wisest to deal with the problems now.

I want to comment about AIDS and drug treatment. Everybody is terrified about the implications of the growing threat of AIDS. They say that now we have to deal with the drug dealers because through IV drug use the disease could spread to "us." This fear may make more treatment money available because we are going to be able to show that drug treatment is AIDS prevention. But in social terms, I do not see much difference between the progressive terminal disease of drug addiction and the progressive terminal disease of AIDS. Now that IV drug users constitute the majority of new cases of AIDS, we must address both problems together.

How do we approach this overwhelming problem of drug use? We must start by providing health care. Part of the problem is that we do not provide payment for such treatment. That is why I am now pushing for mandated in-patient alcoholism insurance and eventually drug abuse insurance. To a significant extent the medical field and the hospitals will provide the treatment when it is funded. Unless drug abuse treatment is funded it will not be delivered.

We also have to coordinate our criminal justice system with treatment; otherwise we are just spinning our wheels. We have got to take the time to design a reasonable policy for legal and illegal drugs. We have reached accommodation and social consensus with our legal drugs. Nicotine, for example, is a very addictive drug which is also destructive to the victim, yet we do not limit its use. (I am thinking of my wife, who just cannot give up smoking. We have a 12-month old baby, for whom we waited very long. During her pregnancy she would not have even a drop of alcohol because it might damage the fetus. And yet, she had a difficult time with the nicotine.) Most addiction in the United States is to alcohol and prescription drugs. We are just beginning to tackle the abuse of prescription drugs and already doctors are objecting strongly to our newly instituted system of triplicate forms. A well-designed drug policy will have to address all of the varieties of drug abuse.

We must also consider the issue of decriminalization. In New York City there is a virtually defacto legalization of marijuana. Only those who are caught selling or smoking on the subway or in some other inappropriate place are arrested. The question of legalization becomes difficult with the heavier drugs: cocaine, heroin, crack, or the latest designer drugs to come along. Different kinds of drugs will be popular in different areas; the most popular drug in Chicago may not be popular in Washington or New York. I call all of these extremely destructive drugs the drugs of death. As the Assemblyman of Northern Manhattan and the Chairman of the Assembly Committee on Alcoholism and Drug Abuse, I am not about to support legalization of death.

We may, however, want to look seriously at decriminalization for personal possession. Anyone picked up drunk may be thrown in a drunk tank, or in some areas, you might get some treatment. The same would not be true if you were loaded on illegal drugs rather than the legal drug, alcohol. For alcohol, only about 10% of our population is at risk for becoming addicted—yet 15% of the drinking population consumes 75% of the alcohol consumed in the state, while the other 85% of drinkers consume under 25%. This 85% do not have a problem with alcohol and want to be able to continue their drinking. We have controlled access to alcohol by taxing it. We have used a tax policy as a means of reducing consumption, and taxing works. A $1 per gallon increase in the price of spirits of over 24% alcohol by volume significantly decreases the number of cases of liver cirrhosis in the following year. Prohibition was a great experiment which everybody now says did not work. But, from a medical and public health standpoint, prohibition worked. We had significant reductions in all of the measures of excessive alcohol consumption. Many people simply obeyed the law, and the same is probably now the case with illegal drugs. If you legalize hard drugs you open a Pandora's box. I believe the answer to the drug problem is treatment, not legalization.

We have difficulty recognizing alcoholism as a disease. Alcoholic patients come through the emergency room and hospital where little or no treatment is offered. Often the patient feels that if the doctor does not recognize the disease, alcoholism cannot be a disease. We now have 11 hospitals in the State of New York that give inpatients the Michigan Alcohol Screening Test, which identifies potential alcohol abuse through consideration of a number of medical problems. Medical doctors should be able to recognize patients with alcoholism and substance abuse and refer them for treatment. Most of our addicted people are still holding down jobs, and we do not see them until the tail end of addiction, after it has done its damage. Not every substance abuser is somebody on the streets taking crack.

It is ironic that addiction is the only disease for which you are not treated when you show up in an emergency room unless you happen to be near death. It is only at the end of the entire process that unlimited medical care goes rushing in—very, very late. There has to be a better way of delivering health care to the substance-abusing population, and there definitely has got to be a better way of dealing with addiction.

# Drugs and the Social Body

Patricia S. Mann, Ph.D.

As I began to prepare for this program, I was struck by the enduring quality of some of our fundamental attitudes toward the use of drugs. In Plato's *Laws,* for example, the Athenian, who presumably speaks for Plato, defends practices of "convivial drinking" against their outright condemnation, as urged by the Spartan Megillus. Plato's Athenian states that wine drinking should be tolerated as part of a more general "training in self-command," explaining that a certain "indulgence in pleasures" is necessary "as a means towards mastery of them." He cautions, however, that "if the practice is treated as mere play, and free license is to be given to any man to drink whenever he pleases," he could no longer vote for allowing any indulgence in the wine cup in such a city (1).

Our concerns with respect to drugs still tend to focus on issues of moderation and excess. There is a fear in our society that excess has a dangerous edge. The competitive work ethic is no respecter of moderation in its mandates for individual effort or in its allowances for the individual pursuit of pleasure and recreation as earned rewards at the end of the day. There is a consequent nostalgia for a time in which social authority was more evident within structures of everyday life, directing the individual toward meaningful rituals of civic virtue and moderation, in the use of drugs, as in other social practices.

Within our society there is a presumption that the government must wait until inappropriate behavior has generated social ills of a certain magnitude before it can intervene to attempt to restore public well-being. The sorry condition of the current governmental effort to control the use of cocaine and heroin can be attributed to this inefficient belatedness of political authority within liberalism. Many people have argued that the government needs more authority today to deal with a situation in which a dearth of public regulation has encouraged drug trafficking and drug abuse. Polls say that 90% of Americans think the government should be doing more to control the use of hard drugs.

Certainly either more, or less. I think the standard framework for public policy analysis, suggested by John Stuart Mill 100 years ago, may provide sufficient grounds for conscientiously evaluating and rejecting the current policy of drug regulation, and perhaps for supporting the legalization of drugs such as cocaine and heroin, as well. Yet the conceptual limitations of Mill's 19th-century theory make it unable to address the contemporary concerns with drug use and abuse which now form the basis for popular sentiments against legalization.

I want to briefly outline the Millean framework of analysis, since it is the foundation of so many arguments for and against policies of drug interdiction. I then attempt to show how the present social meanings of hard drug use push us beyond Mill's 19th-century notion of an opposition between individual freedom and governmental intervention in individual lives. A satisfactory response to the drug problem will only be achieved, I believe, through a rethinking of our concepts both of individual freedom and of the governmental responsibilities that result from a public commitment to such individual freedom.

**The Millean Framework.** According to the Utilitarian theory of social reform expounded by Jeremy Bentham and James Mill, governmental policies of social reform could and should be undertaken whenever they fulfill the principle of "promoting the greatest happiness of the greatest number." In *On Liberty,* John Stuart Mill reacts

---

Adapted from the author's presentation at the New York Regional Conference on Medical Ethics at the Mount Sinai Medical Center on March 3, 1989. From the Department of Philosophy, The City College of New York, CUNY, New York. Address reprint requests to the author, 120 East 4th Street, 3A, New York, NY 10003.

indignantly to the potentially unlimited governmental intrusion into private lives which an unqualified commitment to his father's Utilitarian principles might justify. With his "harms principle," he asserts what he deems a fundamental threshold condition for governmental authority to intervene in private lives within democracy. If one person is harming another person in some way, and thus causing what he called a "private harm," or if the nation is threatened with some social ill, such as a foreign war, that constitutes what he called a "public harm," under these conditions and only under these conditions does the government have warrant to consider intervening in private lives.

The harms principle involves a refinement rather than a repudiation of Utilitarianism, however. Once a harm is determined to exist, the second question is whether the apparently necessary act of governmental interdiction will actually promote greater well-being, or whether instead the act of intervention will generate social problems greater than those it was meant to resolve. Policies cannot be justified, even in the face of great social ills, if they come at too large a social cost (2).

**The Social Cost of Interdiction.** After reading a number of articles delineating the current status of the war against illegal drugs, I find it hard not to conclude that the cost of this interdiction effort is too high. On the one hand, it is generally assumed that drug abuse and addiction will increase if hard drugs are legalized. I am willing to concede that such addiction constitutes a public harm, justifying and even requiring some sort of social response. On the other hand, in making heroin and cocaine illegal, and in attempting to prohibit the use of these drugs, we have created a war.

It is not so much a modern ideological war as it is a Hobbesian war of every man against every other man, whether in the pursuit of drug profits or in the pursuit of drug profiteers, the latter aim often collapsing into the former. Gunshot wounds received in the urban war zones of the illegal drug trade alone account for $1 billion annually (85% of it taxpayer money) of treatment costs, according to a *New York Times* article. When medical doctors rather than political activists begin to draw analogies with Vietnam, as in this article, one begins to take notice (3). When police agents in the narcotics units responsible for the major drug busts in New York City say they have no illusions about the possibility of exerting control over drugs once the drugs have entered the country, at the same time that U.S. agents in Colombia are saying they have no illusion about their ability to stem the flow of drugs into the United States, one wonders what interdiction has come to mean (4). Since crack sells on New York City streets for as little as $3.00, and in a recent national survey, 55% of high-school seniors said that cocaine is easy to procure, support for the war on drugs cannot be related to its effect on supply, or to any other obviously measurable benefit (5). We now have "criminalized free access," to use two terms bandied about in what, I think, is the correct juxtaposition.

There are many reasons—economic reasons being high on the list—why legalization may not happen. But what interests me, as a social philosopher, is popular perception of the issue. The analysis above shows, I think, that the overwhelming popular support for legal sanctions against hard drugs must be understood in moral or spiritual terms rather than in pragmatic utilitarian terms. That is neither surprising nor lamentable. Yet drugs are everywhere in our culture, and our reasons for accepting some and rejecting others need to be better understood.

**The Spectrum of Drug Uses.** Consider the spectrum of drug uses, from "good" to "very bad illegal."

"Good" uses are on one end of the spectrum. Drug taking is generally approved when it is for the purpose of restoring or maintaining the normal functions of the individual bodily self. Whether the drugs are available only for restricted use as administered by a member of the medical establishment, or generally available as vitamins or cold pills for self-application, drug use for the treatment of ills is uncontroversial. We worry that particular drugs are not doing what they are meant to do, some people being more skeptical than others about unproven drugs. We do not often question more generally the medicinal goal of seeking to cure individual bodily ills by means of drugs.

"Anomalous" uses are next on the spectrum, encompassing use of a number of relatively new drugs and uses of drugs that have a variety of capabilities going beyond traditional medicinal uses: drugs for contraception, for fertility, for athletic endurance, strength, speed, drugs enhancing our bodily capacities, enabling us to do things we could not otherwise do—some accepted, others rejected. Some of those accepted most readily would seem to pose the greatest threat to fundamental social structures.

"Socially accepted bad" uses are next, involving drugs known to be physically or psychologically harmful to the individual user but continue to be consumed legally and uncontroversially. Cigarettes and alcohol, the major drugs in this category, are associated with 450,000 deaths annually. Popular acceptance of such dangerous, destructively used drugs might seem hard to understand, particularly in conjunction with the popular repudiation of other drugs.

Alcohol and cigarettes, however, have a respected niche in the normative social dynamic of individual responsibility and freedom. As adults enmeshed in liberal social structures today, we accept as basic to individual self-respect the responsibility of finding employment sufficient to maintain ourselves and any dependents we may have. As reward for this responsible behavior, we expect that our earnings will enable us to partake freely of an array of pleasurable and recreational goods and experiences. In this society, our sense of individual freedom is fundamentally bound up with our right to buy, possess, consume, or participate in whatever the market makes available.

One strength of J. S. Mill's harms principle lies in its ability to distinguish between harms directed at others and harms directed at oneself. At the limit, our personal freedom includes the freedom even to engage in activities and consume things which may bring harm upon ourselves, so long as we are not also harming another. Alcohol and cigarettes are consumed in this context. We may remonstrate and plead with our friends and loved ones should they drink or smoke excessively; yet we know that the choice is finally theirs to make. One might say that the significance of our freedom is demonstrated, at the limit, by our freedom to engage even in self-destructive excess.

"Very bad illegal" uses are farther along the spectrum. Heroin, cocaine, and related compounds are central to what one might label an "anticulture of addiction, crime, and poverty." Within this typical urban anticulture, lives are defined as much by the lack of middle-class employment and gratifications as by the more visible relations of drug addiction and crime. Yet the dramatic penumbra of deviant acts which seem to surround lives of addiction are what we associate with the use of cocaine and heroin. Of course, these drugs are sometimes used by those within middle-class culture. They are commodities of titillating risk and excitement, expressions of the freedom to consume at one of its antisocial limits. But we do not normally associate very bad illegal uses of drugs with the middle-class dynamic of responsibility and freedom. Instead we associate such drug usage with the anticulture rituals of unemployment, basic physical neediness, and the dominance of activities related to illegal drug use and commerce. Drugs have, to a certain degree, transformed the quality of urban underclass marginality in our society: relative passivity has given way to violent forms of nonparticipation.

**Opposition to Legalization: The Notion of Cultural Decline by Addiction.** In the social logic of opposition to the legalization of drugs, the first, and probably dominant, form of popular opposition comes from those who fear a moral and cultural slide from "socially accepted bad" use and abuse of drugs to "very bad illegal" use and addiction.

One might label this the "cultural decline by addiction" thesis. The idea is that a tremendous abuse of legal drugs such as alcohol, tranquilizers, and cigarettes already occurs. In a consumer society, where individual happiness is presumed to lie in the goods and experience one can buy, if cocaine and heroin were legalized, these too would become substances of middle-class experimentation. But these drugs are said to be so much more addictive and so much more destructive of one's ability to function that all vision foresees the decline of middle-class culture into a vast suburban commune of nonworking zombies.

Paradoxically, it is those of us who have been most critical of the materialism and debased values of middle-class culture who are most vulnerable to these fears of the loss of this culture. If people really are bored with their jobs, cynical about the old themes of individual achievement and civic pride, burned out by the constant struggle to earn enough money to finally buy a sense of peace and happiness, what defenses do they have against the momentary allures and long-term disabilities associated with crack? Legalization of cocaine and heroin would open a floodgate on personal acts of escape, and the weak moral fiber of our society would give way, finally.

I think such a vision is unwarranted. It relies on the mistaken assumption that the excesses associated with the middle-class and working-class consumption of alcohol and other drugs lie on a continuum with the excesses associated with the use and abuse of drugs by a contemporary urban underclass. In fact, as Plato emphasized in the dialogue between the Spartan and the Athenian, it is the context of the sipping of the wine, or of the taking of cocaine, which makes all the difference. Alcohol and cigarettes are now used, and abused, in the context of the social dynamic of middle-class responsibility and freedom. The drug excesses of a John Tower or a Kitty Dukakis or a colleague are sad but also highly delimited excesses, committed within the bounds of liberal freedom. They are participatory excesses of individuals who are in little danger of falling more than momentarily off the path of individual performance and achievement valued by our society.

If legalized, heroin and cocaine will be used by a greater number of people in this context, as well. It may be that when consumed by someone firmly within the culture of daily work responsibilities and rewards, crack is not as addictive as it appears when used by those outside this culture and its structures. On the other hand, if it is likely to cause addiction and render one dysfunctional, it will not fit into the pattern of participatory excess, and will simply not become popular. Only 5% of the high school seniors who said cocaine was easy to get hold of had actually tried it; given the attitude of teenagers toward authority, I hardly think its illegal status was responsible for their restraint (5).

It makes little sense to construe the real problems of our contemporary culture, problems of work, family, schooling, media, as implying a middle-class propensity for addiction and dropping out. I am not sure that these complex grounds for cultural turmoil and individual psychological strain imply anything about the use of drugs. An increasing focus on individual health has been responsible for a decline in both cigarette smoking and alcohol drinking in the last several years, in spite of the posited cultural crisis. The assumption of a continuum between middle-class excess and the nonparticipatory excesses of addiction experienced by the poor and marginal underclass is simply an illusion born of an inadequate understanding of the social dynamics underlying each form of drug usage. I am questioning the need, as well as the efficacy, of criminalization as a way of making heroin and cocaine "look really bad," to borrow a phrase sometimes used by Dr. Robert Millman.

**Opposition to Legalization: Specter of the Drugged Underclass.** The second group of people who oppose legalization have a more serious set of concerns, I think. One might label this the "specter of the drugged underclass" thesis. The fear here is that if drug use is decriminalized, addiction will be treated as purely a medical problem, and the social pathologies which foster nonparticipatory lives of drug abuse will remain unaddressed. With decriminalization of the drug problem, middle-class concern for the urban underclass will abate, and former policies of neglect will return. Since medicalization is proceeding apace, alongside present interdiction policies, this fear is relevant in the present social context, as well.

The opposition of black political leaders, such as Jesse Jackson and Charles Rangel, to the legalization of hard drugs was puzzling to me, initially. After all, it is precisely the black and minority urban communities which are most devastated by the violence, corruption, and profiteering of the trafficking in drugs. The various drug wars occur on their front steps, and it is they and their children who are caught in the crossfire. Legalization would remove the war, and the violence, and the seduction of criminal drug activities for profit. Black leaders such as Baltimore's Mayor Schmoke do support legalization on these grounds.

Yet with the decriminalization of addiction, further policy issues arise. Violence alleviated, the economic and social marginality of the urban underclass will remain, as will problems of drug addiction. The drug abuse associated with lives of nonparticipation will not simply disappear. From the perspective of the government, drug addiction will be transformed from a law enforcement problem into a public health problem. But the immediately sanguine connotations of such a transformation are not necessarily warranted. Will addicts be cured, or peaceably maintained? Presumably the same dynamics of economic and cultural nonparticipation which led individuals to succumb to lives of addiction when these drugs were illegal will continue to lead to new cases of addiction when there is legal access to these drugs. The difference will lie in the fact that the addict is no longer defined as a criminal, or impelled to commit crimes to support his or her habit. For this reason, the basic passivity of a drugged existence will become more evident.

The issue is what will replace the anticulture of drugs, poverty and crime once legalization drains the malign blood of illegal drug trafficking and profits from urban poor neighborhoods. As criminal activities, drug use and trafficking focus our concern on the urban underclass. What would be the nature of our concern for this population once addiction became merely a public health issue? Sociologist William Julius Wilson has argued that social and economic dislocations of the last 30 years have, quite literally, isolated a large population of "truly disadvantaged" in the inner cities, without physical access to the jobs which anchor middle-class lives of responsibility and freedom. Only a variety of concrete economic and social interventions by the government will enable this population to begin to participate in society productively. Yet with the dismantling of '60s Great Society programs there has been what Wilson terms a "demise of the liberal perspective on the ghetto underclass," and conservative pessimism about curing the problems of inner-city nonparticipation has again become dominant (6).

Given the assumption of an intractable "culture of poverty," one can imagine a "kinder and gentler" policy of liberal drug dispensation, resulting in the quite literal pacification of previously angry and troublesome urban masses. Drug maintenance programs might become the pragmatic '90s way of coming to terms with the underclass marginality which the Great Society programs of the '60s failed to overcome. No one would intentionally devise a strategy of drugging and medically maintaining the social body of the urban underclass. It would just happen, for lack of positive programs to overcome the unemployment and social alienation which make drug addiction a characteristic dynamic of contemporary underclass existence. In the recent call by Daniel Patrick Moynihan for a "methadone clone" of cocaine, one can hear the seductiveness of a medicalized response to the complex social problems of the urban underclass. Of course, as Moynihan says, "once a patient is taking methadone the various 'talking' therapies have a chance" (7). A chance, indeed. But how great a chance; and a chance for what? I think there is a very real social choice involved with the medicalization of addiction, whether it is seen as a humane alternative or as a supplement of criminalization.

**Conclusion.** In light of this analysis, I conclude by focusing on what promises to be one of the most difficult ethical issues facing the medical establishment in the context of present and future drug policies. Whether or not hard drugs are legalized, an increasing medicalization of the problem of addiction is on the social agenda, the AIDS linkage with addiction being one primary factor encouraging this trend.

On the one hand, drug addiction has traditionally been seen as a problem of individual excess, a moral failure, à la Plato and various religious stances. And it is a problem of individual intemperance. Yet it is increasingly evident that we cannot combat addiction, in its most characteristic contemporary urban underclass forms, by placing further legal restrictions on individuals. This particular phenomenon of individual excess expresses an underlying social and cultural deficit which must be addressed as its primary cause. An individual here and there will recover from addiction through religious conversion or sheer willpower. Yet the underclass's lack of participation in productive, rewarding modes of life, and its consequent vulnerability to addiction, will not be cured by any mere spiritual strengthening process, if William Julius Wilson's structural analysis of demographic causes of nonparticipation of the urban underclass is even partially correct.

On the other hand, a modern scientific and humanistic orientation encourages us to view drug addiction as an affliction of the individual physical body. We analyze and compare the addictiveness of cocaine when administered to rats and primates to its addictiveness in human hands. Yet the relevant comparison between animals and human beings cannot be a purely physical one; the complexity of human motivations must be contrasted with any presumption of a primarily physical basis for animal behavior.

Human beings take drugs in response to the social meanings associated with the physical effects of the drugs. The social meaning of cocaine addiction for members of the urban underclass must be addressed. Drugs like methadone cannot restore functional personhood outside of a context of meaningful social goals.

There is unfortunately a tendency to oscillate between these two theories of addiction: It is either a problem of the individual will, or a problem of the individual body. Yet I have tried to show that the drug excesses of urban nonparticipation can only be understood as pathologies of the social body. (Insofar as we want to combat the alcoholic and cigarette excesses of working-class and middle-class individuals, these too are probably better dealt with as harmful rituals of the social body.) This means that the propensity toward such excesses can be successfully treated only if seen as problems of the social body, and not as spiritual or medical problems of individuals.

I can now explain what I meant when I said that John Stuart Mill's 19th-century notion of a stark opposition between the free individual, and governmental intervention which reduces individual freedom, would prove inadequate. It has become increasingly evident in our century that what Mill thought of as an inborn quality of individual freedom is no more than an inborn disposition or potential. This disposition toward responsible freedom must be nurtured and developed, just as a disposition toward morality must be developed through the proper training and rewards. There are social structural reasons why various traditional moral dispositions are underdeveloped in many of us in this society. Kindness, candor, fairness—all these may be slighted in the fast, frayed transactions of competitive professional lives like ours. Different social structural reasons explain why a particular population in this country has a lesser-developed disposition toward responsible freedom, or toward the activities we associate with such freedom. The social excesses and deficits associated with this underclass are likely to persist until we find ways to combat its present social and economic marginalization.

If hard drugs were made legal, the war against these drugs would no longer obscure or distract from the structural social and economic features of the underclass situation which make this population vulnerable to drug-addicted lives. Social policies to increase employment opportunities and other forms of productive social participation might well become a higher social priority. This will not be a rapid or an easy process, and the temptation, to fall back on treating addiction as merely a public health problem of individual bodies, will remain.

Medical doctors are the last line of defense against such a medical shortcircuiting of the difficult process of social restructuring that is necessary if the underclass is to be offered any real opportunity for productive social participation. Can medical doctors draw a principled line between the momentary treatment of withdrawal symptoms, and longer-term treatment which involves physical maintenance of the addict? Can the medical profession insist that “very bad” use of drugs does not become “good” use merely through being administered in a medical context?

These questions need to be addressed. However they are answered, the existence of such questions places the medical profession in a morally pivotal position. Contemporary medical structures dispose doctors to focus all their attention on the individual physical body placed in their care; the moral and social implications of the act of physical treatment tend to be obscured, and all too easily ignored. What will you do?

### References

1. Plato. Laws, book 1, 637d, and book 2, 673e-674. In: Hamilton E, Cairns H, eds. The collected dialogues of Plato: Princeton, NJ: Princeton University Press, 1971.
2. Mill JS. On liberty. New York: Penguin Books, 1984:68, 119, 141. See all of chap. 4 for relevant policy analysis.
3. Gross J. Epidemic in urban hospitals: wounds from assault rifles. New York Times, Feb. 21, 1989.
4. Kifner J. Soldiers in the crack war: fighting street by street. New York Times, Feb. 22, 1989. Massing M. The war on cocaine. New York Review of Books, Dec. 22, 1988.
5. Berke RI. Student survey detects decline in use of crack. New York Times, March 1, 1989.
6. Wilson WJ. The truly disadvantaged. Chicago: University of Chicago Press, 1987:5.
7. Moynihan DP. Yes, we do need a methadone clone. Op-Ed article, New York Times, Feb. 26, 1989.

# The Physician as a Servant of Public Policy

Barry Stimmel, M.D.

The recent attention given to ethical issues in medicine, often accompanied by a public view differing from that traditionally seen in medicine, has made some physicians less than enthusiastic about entering the public arena. Some physicians' anxieties are such that their alarm at being surrounded by ethicists is exceeded only by the prospect of a room full of litigators. However, physicians have always looked critically at the "hard" issues in medicine that they face routinely in their daily practice. Whether physicians can or should act as servants of public policy with respect to substance abuse, considering the crisis occurring in many areas of our country, is a relevant and far from theoretical question. However, before I discuss the way physicians have reacted to this issue, a more important question must be addressed: Is there a coherent public policy concerning substance abuse clearly stated and actively pursued by the public?

## A Coherent Public Policy?

To some it may belabor the obvious to question whether we have established a consensus with respect to substance abuse. In 1988 Congress, in an extremely well-publicized effort, passed the Omnibus Drug Bill, PL 100-690, a piece of comprehensive legislation directed toward curbing both the supply of illicit drugs and the public's demand for them, as well as providing greatly expanded treatment facilities for those using such substances.

The message conveyed was clear and unequivocal. An additional $2.8 billion over the 1988 level of $3.5 billion was allocated for this effort. Almost all legislators were ecstatic in their praise of this bill and their commitment to ending the "scourge of drug abuse." However, to step back from the rhetoric and look at the facts concerning this public policy makes one less than enthusiastic. Some random observations of a more cynical bent might illuminate issues that the bill does not address.

**Curbing Supply.** We in the United States have a long history of supporting governments whose officials have been actively profiting from the distribution of illicit substances in our country. In almost every instance where the "greater political good" was seen to be served by supporting a particular government, such support has been given. Our aid was given irrespective of knowledge of involvement by government officials (including, at times, even the government "leader") in aiding or abetting the transport of opium, cocaine, and marijuana into this country. In fact, this policy was unambiguously declared by the State Department:

> Political and economic instability in drug-producing areas around the world has resulted in the subordination of our drug control agenda to other pressing concerns (1).

This policy was also bluntly but no less eloquently expressed by Nathan, a 28-year-old black heroin user from Washington, DC, whose only insight into the political scene is based on the fact that he lives in our nation's capital, which has one of the worst drug problems in the country. He said,

---

Adapted from the author's presentation at the New York Regional Conference on Medical Ethics at the Mount Sinai Medical Center on March 3, 1989. From the Department of Medicine and of Medical Education, Mount Sinai School of Medicine (CUNY), New York. Address reprint requests to the author, Dean for Academic Affairs, Box 1255, Mount Sinai School of Medicine, One Gustave L. Levy Place, New York, NY 10029.

> There was a time in my life when I wasn't usin' any drug, when I tried to have the middle-class American dream. Have yourself a little piece of property, send your kids to a nice school. The "All American Dream." I had those dreams too. . . . I've got those dreams. But that's all it has been is dreams and visions because you can never reach the type of success that they can. . . . Dream? What kind of dreams can a Black man have in this country? They control those things. Some of us are able to deal and accept that and those are the ones that don't have to use drugs. Politics is run just like everything else in this country. Those with money get what they want and those without don't get shit. . . . So whatever happens, happens, and we have to accept it. But the main thing is, what are we doin' to control the importation of heroin into this country; the people who are makin' the big money off of this; what are we doin' to the people in this country? Those are the big questions. And, we are avoidin' those issues. We're only fuckin' with the small man. We messin' with the low man on the totem pole. When the shit rolls downhill, you don't start at the bottom of the hill . . . you have to go to the top of the hill. That's where the decisions are made. You go to the top and you start choppin' heads at that end. . . . There would not be no drug problem if there was no drugs in this country. What are we gonna do to the people bringing drugs into this country? That's where you start. Don't start with me or Joe Blow. Me and Joe Blow between the both of us ain't got twenty dollars, we can't even cop a quarter. We got to go out and steal money to get some. Don't start with us. Why give me thirty or forty years just for usin' some? You get the man that gets the money from me usin', get him. Do somethin' to him. And, that's when we'll all be better. Until we decide that we gonna do that it's just gonna be a vicious circle. (2)

Many, like Nathan, feel that our policy toward controlling the entry of illicit substances is both far from coherent and inadequately implemented.

Our track record, with respect to licit substances, also leaves much to be desired. Although the surgeon general is quite appropriately against cigarettes, their sales continue unabated, and the subsidies provided to tobacco farmers are still impressive. Our extremely stringent Food and Drug Administration, which has been criticized for delay in the approval of beneficial drugs, has been noticeably lax with respect to regulation of mood-altering substances. Propoxyphene (Darvon), initially promoted as a non-dependency-producing drug, is still available today without any restrictions other than a physician's prescription. In fact, it is an ester of methadone that has been used in maintenance for heroin addicts and can be easily abused, producing considerable dependence and withdrawal when taken in high doses. Pentazocine (Talwin), touted as an effective pain reliever, is also uncontrolled. It is still an extremely "hot drug" on the street: under the label "T's & Blues" in some localities, it is injected to produce a high identical to that of heroin. A whole new generation of pain killers, the narcotic analgesic-antagonists, has been produced and allowed on the market without adequate regulation. Although these drugs are promoted as having an extremely low risk of dependency, in experimental trials heroin addicts have described the high achieved from some of them to be similar to and, in some cases, more pleasurable than that of heroin.

As demonstrated by the decreasing street price of cocaine, we remain fairly inept at controlling the supply of illicit drugs. Furthermore, we foster the production of mood-altering substances, such as tobacco. Nevertheless public policy, especially in the state of New York, has moved into physician prescribing, an area that it can control with somewhat greater ease. The attempt to prevent physicians from prescribing medically unwarranted benzodiazepines is certainly laudable. However, to accomplish this end by requiring physicians to write triplicate prescriptions appears to be somewhat less than effective. Surveys have documented that patients who are prescribed the benzodiazepines for medical reasons consistently tend to use less than the prescribed amount and reduce use over time (3). This model can be contrasted with the pattern of individuals who are intent on abusing such drugs; they visit multiple physicians' offices to obtain limited supplies, or frequent Medicaid clinics where dispensing of such medications is legion.

It is likely that the use of triplicate prescriptions will considerably interfere with the ability of the former group to obtain their needed medications with confidentiality, but will have little effect on the latter group, which has no problem making the many required trips to different physicians' offices for a 30-day supply. The creation in the New York State Health Department of a data bank containing names of persons using these drugs for legitimate medical reasons might also be less than medically desirable and even constitute an infringement on privacy and a limitation on liberty.

**Curbing Demand.** Obviously, diminishing the demand for drugs has great importance in addressing the problem of substance abuse. Contradictions, however, in policies aimed at curbing demand also appear despite "public commitment" to that goal. In order to increase the undesirable consequences for individual drug users, the 1988 comprehensive antidrug act established an array of penalties, including loss of eligibility for federal loans for education and housing. Given that the addict already takes considerable personal risks, including death from overdose, it is unlikely that the inability to obtain a college loan will serve as a deterrent. However, the opportunity to use such programs could be extremely important to those in treatment. Denying such options may well impede the rehabilitative process.

**Education.** Education is, of course, the most potent weapon in the war to diminish drug use. However, of the billions of dollars allocated to this effort by Congress, only $500 million have actually been identified. To this date, it remains unclear to all, including the federal administration, how the remaining funds will be made available. With New York State already under severe financial constraints, it is also difficult to imagine that the state would undertake an increased effort.

**Treatment.** Treatment, of course, is necessary for those who already are drug dependent. Unfortunately, we are not even close to meeting the laudable goal of having treatment available to every person with need, and there is no indication that we will achieve that goal in the near future. Indeed, unlike other illnesses, which are covered by third-party carriers, insurance coverage for substance abuse is almost nonexistent.

No less important, however, is the absence of adequate research to determine which methods of treatment are effective. For example, a great deal of money is directed toward inpatient treatment for alcoholism, when the evidence suggests that many people would benefit equally from treatment in the outpatient setting at one-tenth the cost (4). With a similarly unscientific approach, the federal government has recently announced its intention to relax regulations for establishing methadone maintenance programs. New programs would then provide minimal counseling to large numbers of heroin users (5). This confusion of "treatment on demand" with "drugs on demand" might well be followed by withdrawal of funds for demonstrably effective rehabilitative services.

Despite our war against drugs, little effort has been directed toward those factors associated with the high risk of drug use. These risk factors are, in fact, the reasons why most turn to drugs in the first place: children left without supervision (due in part to inadequate or nonexistent child care), lack of available housing, lack of vocational rehabilitation programs, and absence of viable alternatives to a life of substance abuse. It is no accident, of course, that the worst consequences of drug use are found within the inner city. The existing threadbare fabric of community life in such areas is being totally destroyed by the sale and use of drugs. Yet the assistance requested and required by those most seriously affected does not exist.

### The Role of the Physician

What has been the physician's role as a servant of public policy within this framework? On the surface the physician is certainly as distressed as others by the environment in which substance abuse flourishes, and as committed as others to ending substance abuse. More than others, the physician is aware of the drain on limited medical resources produced by the extensive medical complications associated with both licit and illicit drug use.

The response of physicians to these problems, however, has been and, unfortunately, remains less than adequate. In part this has been due to a view, not entirely incorrect, that substance abuse is really not a medical problem. The reasons why individuals turn to drugs must be addressed if any progress is to be made in eliminating or diminishing the use of mood-altering drugs. But since the reasons for use of such substances are not perceived as medical, the subject is far from prominent in the medical curriculum. Insufficient teaching time is devoted to alcoholism and substance abuse, leaving the physician (as student, resident, and later as attending) ill prepared to identify early those persons who are at risk for future drug dependency and without the knowledge of how to intervene.

However, even when concerned physicians are aware of the early signs of drug abuse and are willing to engage in intervention and counseling, little or no financial reimbursement is allotted for such activity. Despite the recent emphasis on the importance of speaking with and counseling patients who are at risk, because such a process is quite time consuming it is not pursued as often as it should be. Since one may well see three or four patients in the time it takes to counsel one, only a committed and rather financially independent physician devotes time to this activity.

Sadly, alcoholics and drug addicts are not the most attractive people; nor, when using substances, do they view the medical establishment warmly. As a result, physicians tend to shy away from dealing with such patients in other than purely medical terms. This in turn creates a cycle of hostility, which interferes with the initiation of a beneficial patient-physician relationship. Finally, and perhaps most frustrating, those physicians who do care and attempt to intervene find the resources available for patient referral woefully inadequate. The cost of continuing treatment is beyond the financial reach of most of those who need it who are not covered by Medicaid.

On the other hand, the physician is constantly sensitized by accusations of "pushing drugs" and goes to extremes to avoid prescribing those, such as narcotics, customarily perceived as "bad" drugs. This degree of paranoia or "opiophobia"—a term coined by Morgan (6)—has actually resulted in underprescription of analgesic drugs for pain relief to people who clearly need such medication in order to function. Drugs perceived as nonnarcotics, such as pentazocine, propoxyphene, and the antianxiety agents, although equally dependency producing, are prescribed in excess in an attempt to avoid the appearance of "addicting" patients to narcotics.

Faced with these obstacles, the physician tends to address the issue of substance abuse only when it is inescapable. At that point the "medical model" is invoked, often to the detriment of the rehabilitative process.

Considering all I have said here, one may think I am an adherent of the views of the Marquess of Salisbury, who stated, in 1877, "If you believe the doctors, nothing is wholesome; If you believe the theologians, nothing is innocent; If you believe the soldiers, nothing is safe" (7). And one might even add, "If you believe the politicians, everything will be better once they are in office, but nothing is further from the truth."

I believe that adequate public policy on substance abuse can be designed and that the physician can function in good conscience as a servant of that policy. Also, I believe that our efforts toward diminishing inappropriate use of mood-altering drugs can be successful without either imposing draconian penalties on those who seek relief through the use of such substances or making all substances freely available in a competitive marketplace.

What is needed is an accurate perception of the problems surrounding substance abuse, accompanied by appropriate commitment of resources and efforts on the part of all concerned, not merely the physician or the politician. Committing ourselves will allow Nathan's dream of being "community oriented; go to a nice church; do things in your community; make sure the schools are really giving your kids a good education; dress nicely; eat good food; do things . . ." to become a reality for many more people, including those whose lives and families have been destroyed by substance abuse.

## References

1. Rosenthal AM. The giant loophole: how the U.S. circumvents anti-drug law. New York Times, March 3, 1989, A15.
2. Hanson, B, Beschner G, Walters JM, Bovelle E. Life with heroin—voices from the inner city. Lexington, MA: Lexington Books, 1985.
3. Woods JH, Katz JL, Winger, G. Use and abuse of benzodiazepines: issues relevant to prescribing. JAMA. 1988; 260:3476–3480.
4. Hayashida M, Alterman AI, McLellan AT, et al. Comparative effectiveness and costs of inpatient and outpatient detoxification of patients with mild to moderate alcohol withdrawal syndrome. N Engl J Med 1989; 320:358–365.
5. Altman LK. U.S. to ease methadone rules for addicts to combat AIDS. New York Times, March 3, 1989, A1.
6. Morgan JP. American opiophobia: customary underutilization of opioid analgesics. Adv Alcohol Substance Abuse 1985; 5(1/2):163–173.
7. Talbot Gascoyne-Cecil RA. Letter to Lord Lytton, Viceroy of India, June 15, 1877. In: Bartlett J, Beck EM, eds. Familiar quotations. Boston: Little, Brown, 1980.

# Panel Discussions

## Drug Use: Social and Scientific Background

MODERATOR: DANIEL A. MOROS, M.D.

**Daniel A. Moros:** To start, two related questions, which I address initially to Professor Fox. It has been suggested that there are many reasons for making some drugs illegal and permitting relatively free access to others. Why does our society consider some drugs undesirable enough to invoke the law to prevent their use? Also, what would a rational drug policy look like?

**Daniel M. Fox:** It is important to consider each drug individually. A wonderful book was published recently about the introduction of the coffee house into the Islamic world during the fifteenth century, much to the dismay of the ayatollahs. Understanding the process by which the society adopted coffee and the coffee house and ultimately made it a desirable addictive behavior provides one model of how drug use and drug policy emerge over time.

As for a rational drug policy, let me say, exaggeratedly, that as I look at drug policy in the United States over the last hundred years I do not think the word "rational" is useful. In a broad sense every country gets the health policy it desires, and our drug policy, like our larger health policy, comes out of our culture and our ways of approaching social problems. Consider the three approaches—social control, tolerance, and therapy—in the United States; compared with Britain or Sweden, we have chosen to move away from policies of tolerance such as supporting maintenance programs. Seeing that policy as rational or irrational is not the question so much as is understanding the forces that led us to make our particular choices at each particular point in time.

**Moros:** That may well be true, but the ultimate purpose of our inquiry is presumably to help those who make public policy make it better. Dr. Millman and Dr. Galanter, what should we be doing?

**Marc Galanter:** Unpredictability has been a characteristic of the history of federal efforts to manage substance abuse. For example, because no one ten years ago anticipated crack, any intelligent discussion of policy in 1979 would not address our most pressing concerns in 1989. Certain measures that might be useful, like needle distribution, cannot be implemented in today's political climate. But social and political perspectives could change in a few years. The legalization of marijuana, which seemed on the way ten years ago, is now as dead as pro-child-abuse legislation. You can bet that patterns of drug use will change in the next five years. Those changes may make anything that today looks like a viable policy inappropriate. I do believe that humane treatment for the substance abuser is something doctors are capable of. It will always be appropriate.

**Robert B. Millman:** On the question of what a rational drug policy would look like: If we could pick which drugs our society should abuse, I think we would choose marijuana and we might even choose heroin; we would make alcohol illegal because it is too dangerous and too toxic, and I think cocaine also is clearly too dangerous and too toxic to be readily available. About cocaine I find myself thinking that we must become increasingly draconian.

We must try, in some way, to change society's attitudes: we must make drug use look unattractive. Use will decline only as society perceives drug abusers as ugly, or dumb, or jerks. There are a number of ways of accomplishing this goal. In Singapore the first thing you see when you get off the plane is a big sign saying "Death to the Drug

---

Adapted from the panel discussions at the New York Regional Conference on Medical Ethics at the Mount Sinai Medical Center on March 3, 1989. Address reprint requests to Rosamond Rhodes, Coordinator of Medical Ethics Teaching, Box 1193, Mount Sinai Medical Center, One Gustave L. Levy Place, New York, NY 10029.

Sellers." You believe they mean it; in fact, in Singapore they do kill drug dealers, and Singapore is "clean." I am not advocating this particular measure, but there are many others. All over Singapore and Malaysia you see posters in which drug addicts and their mini-society are sloppy and ugly, quite different from our romanticization of drug use. I hate the idea that I am becoming so "illiberal," but I do think the government needs to maintain a fairly heavy hand in trying to make drug use look less attractive. Should people be able to use drugs with impunity? I do not know how to deal with them. Many drug users do not bother anybody and do not get into trouble, and maybe they should be allowed to do what they want. But though I hate to agree with Nancy Reagan, they are fueling the drug trade. Certainly I would also argue for humane and more readily available treatment.

**Don C. Des Jarlais:** Since Dr. Millman is uncomfortable with what he is saying, I will take the opportunity to disagree with many of his views. One point on cocaine, the Incas used it for hundreds of years without any major problem. It was not a matter of the drug, but rather the way they were taking it. They were chewing it. The dosage was relatively low and, taken in this fashion, the effects had relatively slow onset. If people smoking crack were chewing cocaine, we would not be worried about a cocaine problem. Rather than simply identifying the most dangerous drugs, we need to look at the way in which the drugs are taken, the mode of administration, and the social setting.

**Millman:** I agree. But what do we do now when we have the drug in its purified form and the technology to deliver it in high dosages?

**Des Jarlais:** One thing to note is that the middle class is using drugs a lot less; cocaine is going out of style quickly in the middle class. What is happening now is very much what happened after the institution of the Harrison Narcotics Act. Then people who had nothing to lose continued using narcotics. Today the emotionally ill and disadvantaged, because they are already so miserable, continue using drugs. Chances are today's drug use will spontaneously decrease, we will develop better methods for treating addicted people, and the epidemic will trail off. I would guess this will happen over the next twenty years, just as the heroin epidemic trailed off as we developed better means of treatment. Our current technologies and treatments need to improve and we probably need programs to get people off the streets. What will happen after that is unpredictable, just as each of these waves of drug use has been unpredictable. But I would not hang too many addicts, because chances are the problem will abate. I do not want to preach to you, but our awkward system of civil liberties is worth protecting, even if it sometimes limits options for policy.

**Millman:** I have no argument against civil liberties, but I think we must continue trying to change the ethos associated with drug use. For the young person with resilience, intelligence, and tremendous frustration living in the inner city, one avenue to money, power, and prestige is through the drug world. I do not want to hurt these hypothetical young persons, I want to help them. I think we have to find new ways to give them power. We have to develop alternatives to drugs and make drug users and drug use look unattractive. I am not for punishment.

**Des Jarlais:** But you are for punishment. If you socially stigmatize drug users, if you say they are bad evil people, they are going to be punished by society.

**Millman:** How about "sad and unfortunate?"

**Des Jarlais:** Then they are going to get absolutely lousy, underfunded treatment programs.

**Tziporah Kasachkoff:** You do have to change the ethos, but I think part of such an effort involves the decriminalization of drugs. People are going to go after what they want. As Robert Millman said, if you have five bucks and you want to go to Michigan and listen to the Grateful Dead, you are going to do it. If you are motivated enough to get those drugs, you are going to get them even at an enormously high price to yourself and to society at large. We have tried criminalization and it does not work. We see this every day in the papers and on the streets. We saw it with the prohibition of alcohol.

I will not be shy about using the word rational. It is not rational to persist in a policy that has proven ineffective. It is not only proven ineffective, it is proven extremely hurtful, both for addicts and to the public at large. I would agree that in decriminalizing drugs you remove much of the social stigma. But legalization does not preclude continuing to control access. Children, of course, will come under the same types of paternalistic legislation that we condone now. Decriminalization does not mean that anyone can simply go to the candy store and buy intoxicating drugs. We regulate methadone. It has not been successful because centers have not been efficiently run and the waiting lists are enormous. However, if we were more serious about rehabilitating people, we could do a lot more. So a rational policy must be triple-pronged—decriminalize, regulate, and rehabilitate.

**Millman:** Alcohol and cigarettes are by far the largest killers. Cigarettes are estimated to cause perhaps 300,000 deaths per year and alcohol another 250,000. Illicit drugs cause something like 6,000 deaths—6,000 compared to almost 600,000. I believe that this tremendous difference has something to do with our policy of prohibition. If we license drugs, I think we will see a marked increase in the number of drug-related deaths.

**Lyons:** What happened to the death rate from alcohol during prohibition?

**Millman:** The number of deaths was cut by much more than half, perhaps by 80%. I would like to be able to say this proves my case. But as prohibition continued, year by year the death rate increased. It never got near what it had been before prohibition, but a percentage of the population apparently started to find their alcohol and die.

**Galanter:** There is a difference between making something illegal that has already been widely accepted within a culture and keeping something illegal that is already illegal. The decriminalization of drugs is an experiment with potentially dangerous consequences because, if it does not work, we might never be able to get back to status quo ante.

**Kasachkoff:** Status quo ante is not what we want. That's where we are now. I am not suggesting that decriminalization will solve the drug problem. It will not. It will solve the crime problem. It will eliminate the involvement of organized crime. We need programs of education and rehabilitation, along with regulation.

**Audience:** What was the British experience when they legalized heroin?

**Galanter:** They never legalized heroin. They permitted its use by prescription. You can prescribe it for pain; it is legal to prescribe it for addiction treatment. The general feeling in Great Britain now is that if you are going to use a prescribed drug you should use methadone because it can be taken orally and it is longer acting. Currently, however, most British treatment is based on a Narcotics Anonymous model. They found that having it available by prescription did not get rid of the black market in heroin, though it probably led to more humane treatment of those addicts who came forward for treatment. You cannot draw any simple conclusion from Britain.

**Millman:** I think the lesson in pharmacology is more dramatic. You could not easily maintain someone on heroin because it is short acting and addicts tended to increase the dose. So many of the doctors precribing heroin for addicts found their patients wanting more. This may be one of the causes for the black market in this drug.

**Des Jarlais:** Shreveport, in the early part of this century, had a very good maintenance program for morphine, a short-acting drug, and had results that we would be happy with in any of our methadone maintenance programs. So it is not simply a matter of pharmacology. Much more may hinge on the social situation. The people in the Shreveport clinic tended to come from stable families, to be employed, and to have good social resources. These seem to be the major determinants of outcome for any type of drug treatment.

**Galanter:** But the notion of replacement or maintenance therapy is useless for cocaine, because there is no proper level of prescription for cocaine.

**Millman:** I can just picture the patient coming to you like Dr. Jekyll and Mr. Hyde—a book, by the way, that came out of cocaine experiences. Dr. Jekyll was a mild-mannered medical scientist and Mr. Hyde was a crazed, psychotic character. Your patient, Mr. Hyde, comes in, crazed and psychotic, and says, "I want my dose."

Cocaine is powerfully reinforcing, and makes the addict psychotic. I cannot see how a maintenance program would work. I was completely for legalization of marijuana ten years ago, but not cocaine. When a drug such as heroin is given to an addict, it quiets him down and makes him feel good. But after repeated doses of cocaine users get psychotic and aggressive and become dangerous to themselves and others. One might make an argument for legalizing drugs that do not produce acute psychiatric or physical damage, but not cocaine.

**Des Jarlais:** I do not want to argue for the legalization of cocaine, because I see huge problems. But I do not think our present system of criminalizing cocaine is keeping cocaine out of the hands of people who have a real genetic, or social, liability for it. I do not know if that many more people would be getting their hands on cocaine if doctors were prescribing it. It is certainly cheap enough on the street.

**Kasachkoff:** Remember, nobody is suggesting free access. Legalization of a drug does not mean free access. It may mean regulation.

**Millman:** What restrictions on access to crack?

**Kasachkoff:** Those are details that need to be worked out. Equating legalization with free and easy access sets up a straw man and is not the suggestion that has been made. Some drugs are now available by medical prescription, and we can create other kinds of qualifications.

**Des Jarlais:** We are about ten years into the current cocaine epidemic. We do not really have enough data yet to know what is going to happen to this epidemic. Previous cocaine epidemics pretty much died down, partly because people recognized the dangers and discomforts associated with the drug, and the number of new recruits simply dropped. However, many worry that the present epidemic may have become so embedded in our social system and that the profits from cocaine may be so great that the epidemic will not abate. We know that the situation has changed rapidly over the past ten years, and it may change rapidly over the next ten. It may get much better. If crack is really as bad as the clinical reports, there may well be fewer people starting to smoke crack five years from now. People learn and communities learn. But the situation could also get worse. Among other things, the effect of law enforcement on crack use may be undermined by the vast profits of a growing industry. We simply do not know what is going to happen.

**Millman:** There is actually an interesting pharmacologic, historical point here. People have difficulty continuing to glamorize drugs because they are so painful, they are so paranoia inducing. But you are also refusing to recognize the power of pharmacology and the special addicting potential of cocaine.

**Kasachkoff:** I think you refuse to recognize the power of profits. The financial rewards, rather than pure pharmacology, may well be the critical driving force of this epidemic; it is simply not clear.

**Millman:** Let me finish my point. Dr. Kalb, in 1935, claimed that the disease associated with analeptic drugs, the "up" drugs, is ultimately so painful that abusers eventually use the "down" drugs to self medicate. And indeed, that is what we are seeing with many of our cocaine-addicted patients. They begin using alcohol and heroin because the paranoia and agitation are too great. The natural history of cocaine use may well involve a spontaneous decrease in the number of new addicts.

**Galanter:** It is not clear to me that this discussion is any different from one that would have taken place amongst us ten years ago in a symposium on the pharmacology and control of drug abuse. Where do ethics and morality come into this?

**Moros:** One of the perspectives of the program in Ethics and the Philosophy of Medicine which is sponsoring this symposium is that the ethical issues we confront in the practice of medicine and structuring of health policy are intimately intertwined with judgments about facts, questions about science. We also recognize the need to understand and identify those reigning ideologies that may direct and misdirect our judgments. This discussion is a prelude to the discussion of public policy, which may be viewed as ethics written in broad strokes. To discuss public policy without this background would leave us as poorly informed as many of the legislators, jurists, and medical scientists who formulated our present policies appear to have been.

The dangerousness of different drugs is central to the question of how much we will permit the state to intrude on an individual's freedom. This relates directly to Dr. Millman's concern about whether or not the harmless drug purchaser and user should be punished in order to advance the ultimate success of a prohibition policy. Dr. Millman also suggests that a policy of cocaine prohibition may be justified because it would protect others from crazed, paranoid addicts, and the state has a duty to protect its citizens. If protection justifies state intervention, estimates of 300,000 deaths a year in the U.S. from cigarette smoking and 250,000 deaths from alcohol use raise the same question. All of these ethical issues are imbedded within our discussion, and many have been explicitly addressed.

**Millman:** Let me defend my comments about draconian measures. I am not suggesting that we adopt a policy of death to the drug abusers. Ultimately most are victims, and many of them are out of control: they are addicts, quite often, because of their disabilities. I do not want to hurt any of them. I view them as patients, not criminals, and I do not want to hang any of them. I am not suggesting increased criminal penalties. But my concern is that we cannot romanticize drug use. Society cannot make the life of a drug addict look appealing. We must maintain criminalization so that drug use and the drug abuser appear unattractive.

**Fox:** It seems to me that in the last few years, we have had more people seeking treatment for substance abuse than ever before. There are reports coming in from large HMOs around the country of increasing expenses related to an increasing demand for treatment. To a great extent that demand stems from the availability of treatment and the availability of financing. Of the 620 mandates on health insurance passed by state legislators around the country in the last 20 years, about half are mental health treatment mandates. All of this fuels the demand for treatment. It is ironic to hear this discussion of criminalization vs. decriminalization focus on the demand side when society, now as never before, accepts a medical approach to addiction.

# Drug Use: Public Policy

MODERATOR: BERNARD BAUMRIN, PhD., J.D.

**Bernard Baumrin:** Should triplicate prescriptions be required?

**Barry Stimmel:** What is really required is an effort by physicians not to prescribe these drugs excessively. I don't think a triplicate form will decrease the use of drugs by people determined to abuse them, because those people will simply visit multiple physicians and obtain multiple prescriptions. What the triplicate prescription requirement will do is make physicians cautious about prescribing for a single patient more medication that he or she thinks absolutely necessary; ultimately, some persons who have a legitimate need for more medicine will suffer.

**Stephen C. Joseph:** The literature is extensive on patients addicted to drugs like Valium, barbiturates, and other hypnotics, often in combination with alcohol, who started off using prescription drugs for some particular medical episode. We really don't know if tightening physicians' prescribing, through what at first glance appears a somewhat silly and make-work process, will reduce the amount of addiction. I think this is no more than a small slice of the problem we have been talking about.

**John Brian Murtaugh:** Two thirds of all addiction is to a combination of alcohol and prescription drugs. We are trying to get a handle on this through some form of regulation. We have now prohibited automatic renewal of benzodiazepine prescriptions. If we make physicians think twice about prescribing because their names are going into a computer, we may actually make physicians sit down and talk to patients instead of just signing a prescription blank.

**Stimmel:** I certainly agree that prescriptions for many drugs should not be automatically renewable, and I am all for physicians talking more to their patients. But I doubt that these new regulations will lead to such a change. Besides, I don't think it is appropriate for a data base to be created in Albany with the names of patients who have filled prescriptions for mood-altering drugs. I don't think that doctors should be wondering whether they may, at some later date, be criticized by the state for using these medications. An additional paper trail that may later be used by investigators if someone should complain about a doctor simply does not create a good climate for medical care.

**Baumrin:** Prof. Mann has said that even if all the programs we have been discussing for direct therapies for drug addiction were implemented, they would be useless because they don't address the problem of our modern, predominantly urban underclass, which uses drugs heavily and in which AIDS is transmitted in the context of drug use.

**Joseph:** I found Prof. Mann's talk both interesting and provocative. However, it seems to me that the one thing you really don't take into account in your analysis is that, unlike Plato's cave, our society has a smoothly functioning machine in the streets. Its purpose is to encourage addiction, and it is staffed by people who are adept, clever, and well financed.

**Patricia S. Mann:** This machine is a profiteering machine. Legalization would primarily address this issue.

**Joseph:** By your own arguments, if you legalize, then you fall into the trap of medicalizing the problem. I wonder if we should not maintain our controls and go after the machine with more effective enforcement?;

**Mann:** Where is this hypothetical "effective enforcement" going to come from? There is little evidence to support the idea that we can limit the world's production of these drugs. And there is no evidence that we can interdict enough material at our border to limit the supply in the streets. In terms of your "ideal" of interdiction, I think it is you, and not I, who is in Plato's cave.

**Murtaugh:** We are having a symposium titled "Morality of drug use, license vs. prohibition," so somebody has got to be for the legalization of drugs. It would be hard to get a politician to take that stand (the present major of Baltimore being the exception) because we are worried about the economics of the drug epidemic. First, a lot of crime is driven by people who need money to purchase drugs. Second, what happens to that money after it's collected? We're talking about billions of dollars. Where is it going? What is it buying? In some Latin American countries, this money has a controlling influence.

But we are not just talking about planeloads of cocaine. We are talking about the problem of addiction. It is not a moral failing to be an alcoholic. It's a progressive and terminal disease and an incredibly expensive one. As a matter of fact, it is now known to be genetically transmitted. Fifteen percent of the population is drinking 75% of the booze we consume. The other 85% of the population does not become addicted. But this is not the case with all drugs. True, many people who use heroin do not get addicted, but crack is very addictive, and the addiction is rapid. Most drugs have a long honeymoon period; not crack. After crack, more designer drugs will appear. This is not simply a matter of poppy fields in foreign lands—we are going to have to deal with synthetic drugs manufactured in our own country. Part of a solution will involve following a medical model of treatment. Because many people will obey the law, by keeping these drugs illegal we will reduce the likelihood of individuals trying drugs. That is important. But at the same time we must also avoid the attitude that because a drug is illegal, the user should be blamed for becoming addicted and therefore should not be treated in the emergency room.

**Stimmel:** I think we are under a misconception. Everybody says we are losing the "war on drugs," but this is not correct. The number of heroin addicts in the United States today is slightly less than it was in 1900. The number of high school students using marijuana is down 30% in the past three years. The reason people give for not using drugs is that they are too harmful. We are now being devastated by cocaine, and the reason is probably its powerful reinforcing effect. It is easy for us, as an upper middle class group, to say, "Let them use cocaine." But the people who live in communities that have been destroyed by cocaine go crazy when you talk about legalization. There is a great risk here.

**Joseph:** Since I see no indication that our society is ready to deal with the social problems at the root of the cocaine epidemic, I share Assemblyman Murtaugh's concern with broadening the problem.

**Mann:** Look at some of the contradictory data we have been given today. First we are told that drugs are readily available. Then we are told that drug use among high school students is decreasing. The implication is that availability is not the critical factor influencing use.

**Joseph:** Last year's poll, quoted by Dr. Stimmel, did not include New York City students.

**Mann:** Nonetheless, availability is not the determining factor. People have been taught to be afraid of drugs. My students are terribly afraid of drugs, even marijuana.

**Tziporah Kasachkoff:** Assemblyman Murtaugh's position that drug abuse is not a moral failing but a disease, coupled with the view that we should, nonetheless, be for criminalization while treating drug addicts as if they were diseased and not criminals, may be a position that a political personality is obligated to take. The statements taken together, however, do not form a coherent picture. The criminal justice system establishes criminal fault, that's what it is all about. If the act is not the individual's fault, but a disease, we have no business criminalizing it. If it is a disease, treat it as a disease. But you cannot have it both ways.

**Joseph:** I believe that the three persons on the panel who favor maintaining drugs as illegal all think in terms of punishing the supplier rather than the user.

**Kasachkoff:** I think we must disambiguate our discussion. When people talk of criminal penalties, they often think in terms of incarcerating the user. Indeed, if we decriminalize use, we would find increased pressure for the decriminalization of sales.

**Murtaugh:** As a politician, I have skirted around the question of decriminalization, and frankly I have not yet made up my own mind. I am certain I don't want to legalize what I call the "killer drugs." On the other hand, I view the person who is addicted as being in need of treatment, not punishment. I have already said that one of the greatest problems we encounter with the criminalization of drug use is the treatment of addicted persons. But we control the sale of alcohol, for example by age, using legal sanctions directed against the seller. A bill coming up in this session of the legislature will make the under-age buyer subject to punishment. I continue to believe that if we legalize drugs, we will see increased usage. Furthermore, I don't quite trust what I hear from the medical-scientific community. Ten years ago we were told how addicting heroin was, and how addiction-free cocaine was. Now the reports are reversed. This makes me, as a lawmaker, cautious about instituting major changes in drug policy.

**Baumrin:** Let us allow Prof. Mann the final remark.

**Mann:** I appreciate the difference between the engaged and disengaged positions, and it is no accident that Tziporah Kasachkoff and myself, the philosophers, are those who argue for legalization. But in arguing for legalization I am also arguing for certain alternative policies. When William Julius Wilson points to the loss of a liberal perspective on the underclass, I take that seriously, and it's within our particular drug-use policies that this loss is getting out.

# *Issues in Medical Ethics*

# Doctors' Dilemmas:

## Bioethical Challenges

PANEL: RICHARD GORLIN, M.D., DANIEL CALLAHAN, PH.D., AND BERNARD BAUMRIN, PH.D., J.D.
MODERATOR: AVI BARBASCH, M.D.

**Avi Barbasch:** Let's begin with a hypothetical case: Our patient is a 67-year-old woman, a widow for the past two years, who recently moved to New York to be closer to her family. She has a history of hypertension which is controlled with diuretics, and a history of intermittent angina relieved with sublingual nitroglycerine.

Dr. Gorlin, her daughter is your patient. She calls you and asks if you would be her mother's doctor as well. You agree, but one week before the mother's scheduled visit she develops severe substernal chest pain and is rushed to the emergency room. You get a call from the emergency room resident, who tells you that she's had chest pain for the past three hours, and there are electrocardiogram changes consistent with an anterior wall myocardial infarction. He wants to know whether he should initiate thrombolytic therapy. Should he?

### Who Initiates Therapy?

**Richard Gorlin:** What are the reasons why he might not? That's the first thing. You've predicated two issues here, I think. The first is, Do I really know this patient? The answer is, No, I don't. I've received a story. There seems to be a reasonable time interval, there seems to be a reasonable electrocardiogram change, and I would think under the circumstances two ethical issues are posed: Either the physician who is seeing the patient should be given permission to make the decision, or, if I can personally see that individual within a matter of minutes and assume the attending physician's responsibility, then I can help to make the decision. But if I don't do that, I think I have to delegate that responsibility to the physician who's got the patient in hand at that moment.

**Barbasch:** She has in her pocketbook a small card from "Elsewhere General" with an electrocardiogram which was essentially normal. So in fact new changes are identified. Those changes, and the pain, are consistent with an acute myocardial infarction. You do run down to the emergency room and you are convinced that it is an acute myocardial infarction. Do we proceed with thrombolytic therapy?

**Gorlin:** I have had recounted to me all of the serious objections or contraindications to the use of thrombolytic therapy as best as we can tell. This is a patient who can tell us about previous head injury, trauma, ulcers, bleeding diathesis, all the rest. If this information is provided and no absolute contraindication turns up, you go ahead.

### Choosing between Agents: Benefits or Costs?

**Barbasch:** Both streptokinase and tissue plasminogen activator are available in the emergency room. How do you decide which one to choose?

**Gorlin:** I think you choose the agent based, first of all, on whether the patient has had recent streptococcal exposure or previous administration of streptokinase. So there's no contraindication to using streptokinase. I think you have to make the decision based on the efficacy of each possible choice. Are they equal? What criteria do you use to say they're equal? After that you consider cost

---

Adapted from the authors' Grand Rounds in Medicine presentation at the Mount Sinai Medical Center January 15, 1991. From the Office of the Chairman (RG), Department of Medicine (AB), Mount Sinai Medical Center, New York; the Office of the Director, The Hastings Center, Hastings, NY (DC); and the Department of Philosophy, The Graduate Center, CUNY (BB). Address reprint requests to the moderator, Assistant Clinical Professor of Medicine (Medical Oncology), Mount Sinai School of Medicine, at 920 Park Avenue, New York, NY 10028.

analysis. As far as we know today, in the light of the information available, the effects of these two agents on recurrent myocardial infarction and on mortality are the same, although a truly definitive study has yet to be done. On that basis, I would elect streptokinase, because it's equally effective by today's criteria and it is far less expensive. This decision has some obvious societal obligations.

**Barbasch:** Her son, who you've never met, is sitting in the emergency room. In his hand is the *Wall Street Journal,* which contains an article explaining the latest results on improved survival with the use of TPA and heparin. He points that out to you, and really is determined that you give mom "the best." Does this change your opinion?

**Gorlin:** It doesn't change my opinion, but there's also the issue of practicality. If this individual is standing in the way of my providing some form of thrombolytic therapy for this patient, I think I might be persuaded to do something about it. But I would take a couple of minutes to point out that the *Wall Street Journal* didn't help us much, in fact had been quite wrong on this issue. However, if the individual persists, and says, "Look, I'm going to create difficulties for you if you give my mother streptokinase," I might consult with the patient. You told me she is awake and alert, presumably able to make her own decisions. I might present this to her. So it gets to be a matter of practicality. I don't want to spend an hour in a war. On the other hand, if I can take a few minutes to make a persuasive case, I'll proceed with what I think is right.

**Barbasch:** As the attending physician, not as the chief of medicine, you mentioned your concern about cost. But to whom is this a concern? To the patient? To the institution? To society?

**Gorlin:** I think you can look at it in a variety of ways. We have made the statement that the therapies are equivalent. We're not using the *Wall Street Journal* hype to say that there's a difference between the two, because the currently available studies don't show any significant difference. If that's the situation, I think you have an obligation to say to the patient, "If you have to pay for this, I'd prefer to give you the $80 treatment rather than the $2200 treatment and have you have another heart attack when you get the bill." I think that same obligation is now transferable to whether payment is by Medicare, Medicaid, or any of the third-party carriers. I believe awareness of cost is a major obligation. It's also an obligation to whatever institution you work in. It has nothing to do with being chief of medicine. I'm not the director of the hospital, I don't have to balance the books. But I think you do start with an individual, and you move to the institution and from there to society when you evaluate the issue of cost benefit. I'm sure if I don't say these things, Dan Callahan will.

**Barbasch:** One last question about this to you, Dr. Gorlin. If the patient were your mother, would you still use the streptokinase? Would there be a difference in your approach?

**Gorlin:** I have no problem with streptokinase. I think with the present state of knowledge, it offers certain advantages. We know that TPA is associated with an increased incidence of bleeding into the head. We also know that a prolonged fibrinolytic state is likely to be produced and may prevent the issue of reocclusion of the coronary artery. We didn't get into all the comparative benefits beyond the morbidity-mortality issue. So my mother would get streptokinase.

**Barbasch:** Dr. Callahan, assuming that there are several "best" treatments, how does a physician decide which one to use? Should Dr. Gorlin consider cost in treating the patient?

**Callahan:** I'd want to start thinking through the issue in the following way:

First of all, in comparing treatments—and Dr. Gorlin laid out a nice way of thinking about that—the first question is, What's going to benefit the patient. Then you ask, What are the available treatments? Then you compare those treatments in terms of their efficacy, which Dr. Gorlin did nicely. Then, it seems to me, if cost is a concern, you ask about comparative cost—what will be the consequences institutionally, for the family?

To me the interesting case is what would happen if some slight comparative advantage were found for TPA—not a great or dramatic advantage, but a marginal one. Then a real dilemma arises, it seems to me. But I suppose my question would be, What benefits the patient, what are the comparative costs, what's known about outcome? Then come the larger social questions. But I would only be pushed to those questions having moved through the others first.

In other words, I would not want to start from an economic point of view and superimpose that on the patient, but rather to start with the patient and let the economic issues work as constraints if necessary, but only reluctantly. But nonetheless, if it turns out that two treatments have identical outcomes, the more expensive could be used because a patient's family member wanted it. I could understand why you might do that. But it would seem you'd have a tough time making a case for doing it in the face of some institutional authorities wondering what you are up to. Why do you let yourself be bamboozled by these patients? After all, you ought to be able to stand up to them. I guess I would sum up this way: As much as I'm in favor of considering economics, it seems to me reasonable to consider them only when dragged kicking and screaming into doing so, and only when genuine existing institutional constraints force you to.

**Barbasch:** Let's say that Dr. Gorlin believed the body of literature which demonstrated the benefits of TPA.

**Gorlin:** For example, if I thought the issue of increased patency rate was really important.

**Barbasch:** That's right. There was a true difference between the two agents.

**Gorlin:** There is a difference in the early patency rate which has no effect on mortality. But if I knew fact one but not fact two, how would I be influenced?

**Barbasch:** Would you then advise the use of TPA over streptokinase?

**Callahan:** It seems to me I would ask, How great is the difference? Is it so critical that it's worth spending a great deal of additional money on, and can you make a good case that it is the best medical decision even in the face of resource constraints?

### Cost: The Physician's Responsibility? The Hospital's? Or Society's?

**Barbasch:** Is there an implied contract between the patient and Dr. Gorlin to provide all the care, irrespective of cost?

**Callahan:** I think there's a kind of myth in medicine that the physician does have, with each patient, a contract to do the best for that individual patient. But of course in any real world there are constraints. Even back in the old days, I think 50 years ago, it was not the case that a physician would simply prescribe for asthma a move to Arizona—not for an individual whose income level rendered meaningless that kind of proposal. Resource constraints have always been present, and if you are to allow external pressures to override what is of benefit to the patient, you have to do so on the basis of justice. The problem of justice in health care requires that you do the best for all individual patients. If you help patient A, you may be doing some harm to patient B, and you have to end up balancing.

**Barbasch:** But is that Dr. Gorlin's concern as the patient's primary physician? Does he have to look at the rest of the patients in the hospital, or does he just focus on his patient?

**Callahan:** If the hospital is under economic pressures to remain a viable institution, and it wants to remain one in order to serve its patient load in the best possible way, Dr. Gorlin has to take into consideration the overall welfare of the hospital, which will obviously be impinged on by individual decisions concerning individual patients. I would say that in a hospital you are part of a community. You have to start with that individual patient, but you also have to consider the fact that your colleagues have their patients and inevitably in an era of restraints you're probably going to have to do some balancing. Inevitably, I think this principle of a traditional ethical focus on the individual patient is bound to be compromised in an era of resource restraints. The question is how to do that justly. One way, of course, is to compare, do some technology assessment, look at costs and benefits, and see if you can do that in some reasonable, objective way. The aim, first of all, is patient welfare and, secondly, to justly treat all patients in some roughly equal fashion.

**Gorlin:** Can I just turn that question, Dr. Callahan, one way: Do I do this on my own recognizance? In other words, do I say, this is the decision I'm going to make with restraints? Or is the decision made because of constraints placed by the institution, or our legislature, or some regulator, who says, "Oh no, we're not going to make this available because of cost considerations." If the latter is the situation, then when I see this patient, I have an ethical copout. I'm constrained, I can't offer TPA, it's more expensive, no matter whether it's better—rather than my sitting there making the decision without regard to society.

**Callahan:** It seems to me you would want to reveal to the patient or patient's family any kind of constraints that you operate under. You can explain that certain things are not going to be available because this is a hospital policy, or a social policy, or it's going to be a reimbursement policy, or what have you. If those are real constraints, then you actually can't operate. A limit is placed on you. I think it imperative that some concerted staff discussions about these matters take place so that this kind of burden is shared. It shouldn't fall exclusively on you.

**Gorlin:** I shouldn't have to do it myself.

**Callahan:** Or at least there should be some general guidelines and some notions of what you are each expected to do as members of the staff and what kind of restraint you should set on yourself in the way you treat patients in light of the common pressures on all. Otherwise, you're in an impossible kind of position. You're forced to set a limit on what you do for a patient in what may seem an absolutely arbitrary way.

**Barbasch:** Dr. Baumrin, should we . . .
**Baumrin:** Yes, I'll pick up the question. There's a big picture and a little picture. Gorlin is in the little picture. I don't know why he took this patient on—that was unclear to me from the beginning. He gets a phone call from the daughter that says, this person would be his patient if he makes a decision to have the person as his patient after he sees her, but she has now become his patient because she had a myocardial infarction before he saw her. This seems strange to me. But all right.
**Gorlin:** I took care of the daughter. The daughter says, Would you take care of my mother, who has a heart problem? I said, Sure.
**Baumrin:** Right, I guess that happens.
**Gorlin:** An acute intervention occurred before I had the chance to know her in an elective setting. So I think that much is acceptable.
**Baumrin:** So you've now created a contractual relationship with somebody who doesn't know who you are, so it's not exactly clear to me what kind of agreement you have with her. Knowing what the agreement is will have to wait until she's better, if she ever gets better. I fear she may not get better quickly.

But all right, now you're in there and you're going to act in the best interests of this patient. If we admit, as I think the last discussion is trying to bring us around to, that the TPA is significantly better than streptokinase, you're obligated in your relationship with her to provide that, regardless of what the hospital has decided. The institution's movers and shakers all got together; they decided TPA is too expensive and shouldn't be used. But now Gorlin knows that it's better, so he uses it anyway.

If the physician moves into an institutional setting (that is, takes the institutional view), he or she might just as well move into a national setting, maybe even an international setting. Once you've gone outside the relationship between the physician and the patient, all bets are off as to what you should use. Maybe given the wider view you shouldn't treat this patient at all. It's too expensive, and after all, she's 67, she'll be soon dead. Why give her anything? Was that the question?

### Should Age Restrict Treatment?

**Barbasch:** It's a good question. What would you do if she were 77, or 87?
**Baumrin:** If I had gotten into this situation? Given what I think about these things, I wouldn't consider age a factor except with respect to the person's prognosis. I would choose what I believe to be the best drug, even if the difference were only marginal. I would presumably have a problem with my colleagues, and possibly the hospital administration, about my wasting resources. But it wouldn't matter to me whether she was 57 or 87.
**Barbasch:** So in your view the physician really doesn't have the responsibility to society to reduce the cost of health care?
**Baumrin:** I've thought about that, and I can't see it. I have no deal with society to save it money. I have a deal with particular individuals to make them better, if I can—and I may not be able to. But if I have to worry about larger social issues like that, it seems to me it would get in the way of my providing adequate medical care to anybody—they could even be 17 years old. Maybe I have to assess whether they will use or are using drugs. Do I have to decide whether I'm actually going to provide them with the best care? Do I have to make a moral assessment of every patient, a longevity assessment, and so on? That doesn't seem to be my business as a physician.
**Barbasch:** Dr. Callahan, does age play a role in how we treat patients? Can we restrict treatment based on age? For example, dialysis is restricted by age in England.
**Callahan:** Let me answer this question carefully. I want to make some assumptions. I want to say right now I don't think we have to use age at all at present. We still have enough leeway. But imagine the situation thirty years from now when the number of elderly has doubled and Medicare is running the enormous deficits now being projected. Then we will have to face this issue much more squarely.

The distinction I would want to make is this: it seems to me perfectly appropriate for government reimbursement programs such as Medicare to set limits of some kind on available care, and the question is, Would age be an appropriate, one appropriate, standard to use? My answer would be, Yes, it would, under the following circumstances: first of all, that we have a health care system that has worked hard to make sure that young people have a chance to become old people; secondly, that good long-term and home care is available, that people are not going to be abandoned because of a limit on certain forms of expensive care; and thirdly, that we as a society have mutually agreed on this. It's not something one of us imposes on the other, but we have collectively, through our legislators, decided on that policy.

Keep in mind that we have an age-based program, called Medicare, and we decided that you had to be 65 years old to get in that program; if you're 64, you're out of luck. We might want to say that beyond a certain point, given resource constraints, given the needs of other age groups, given the needs of a lot of young people who are yet to be old people, we're going to have to set a limit. So I would be prepared to see a limit set in that respect.

I would not want to see it done case by case. That would be radically unfair. If it were done, I would want to do it across the board, much the same as for Medicare, age 65 was set as the entry point across the board. We could have set Medicare up for case-by-case eligibility, looking at each and every patient, because some people are old physically much earlier than other people. But we decided a public policy program couldn't be run using case-by-case decisions on who qualifies to be in a program for the elderly. We set an age standard, 65, for access to Medicare. I would do the same thing on the other end, setting a limit to care. I wouldn't do it with individual patients. I would do it by policy, and I would do it across the board.

**Barbasch:** You know, at one time I thought 65 was ancient. Now, I don't think it's old at all. How do you set that cut-off age?

**Callahan:** In one sense, any number you choose is arbitrary. It's arbitrary if we decide to set a speed limit of 55 as distinguished from 54 or 56.

What is arbitrary is saying that age 65 is decisively different from 64 or 66. The point is, in one sense, it will always be somewhat arbitrary, will be within a range. I would want to ask, in the light of resource restrictions: What do we hope to achieve with our health care system for people who are becoming older? How long a life would we want to at least minimally guarantee them on the basis of our available resources? I would then want to ask the question: How long a life in general do most people need to live in order to do most of the things that life enables one to do? Could we come to some rough consensus here? I think it would be difficult. But I suspect an age somewhere in the early 80s might be chosen as a kind of a dividing line. If we had a society that would guarantee everybody a terrific health care system up through ages in the early 80s and only thereafter set some limit, we would have a terrific health care system, far better than the one we have now.

**Barbasch:** How do you respond to those who argue that society would miss out—on at least some of the paintings of Picasso, for example?

**Callahan:** Yes, I think that's a real issue. But another way to look at it is to say that, if you're a Picasso, you're already rich and famous, so we're going to do even better by you. It's a wonderful case of the rich getting richer. If you've made it to age 88, you've got a lot to give, so we're going to keep pouring more money into you because you're a worthy kind of person. Forget the people who are less worthy. But it would seem to me to be unjust to run a society on that basis. So, too, it seems to me, what you really are looking for in a just health care system is a system that guarantees everybody some decent minimum level of care, regardless of their particular contributions. The Picassos of the world by and large can economically get by pretty well. I don't imagine he would have had any trouble buying good health care. He wouldn't have been dependent on Medicare.

**Barbasch:** But Van Gogh might have.

**Callahan:** Van Gogh might have, but again, do you want to use the Medicare program to subsidize genius? Is that its purpose? It seems to me its purpose is to provide a decent level of care for all, not to give special care to people whom we particularly prize.

### Equal Treatment? Or
### Equal Opportunity for Treatment?

**Baumrin:** You're giving me an opportunity to say something I've been dying to say for many years. I don't know what this decent level of health care for all actually comes to. It seems to me there's a basic confusion between equal treatment and equal opportunity for treatment. Medicine is a vast lottery. Take a little town somewhere, pick a state, take a small town in that state. People elect to live there because it's beautiful, it's bucolic, people are nice, they say "Hello," and so on. But if something happens, they can't get a physician there, and maybe they can't even get to the appropriate physician. Other people choose to live in dirty, filthy, dangerous places, and they are two minutes from this great medical institution, or some other great medical institution, where they can get terrific care so they can live another five years in a dirty, filthy, dangerous environment. Now, if we want to promise people equal access to medical care, that's fine—maybe even maximal treatment, given the abilities of the physician involved. But physicians aren't equal; some people know a lot, some people don't know much, some people are able in what they do, and other people are less able.

The most you can give people is an equal opportunity to provide themselves with the services that are available. So limits are already built in.

If you then cut off medical care on the basis of age to those people who put themselves in the way of excellent medical care, it seems that you have now arbitrarily injected into the system something for which there is no justification. The justification for curtailing care can't be said to be equal treatment, because then we should have maximal care for 40-year-old people in small towns. It just makes no sense, the entire gross national product would have to be devoted to health care if we want to provide maximal medical attention and quality services to people who are 40 years old in every little town in the United States.

**Callahan:** Of course, as a realistic matter we understand that if you choose to live in a rural area far from health care resources, we have no societal obligation to provide helicopters and other things to get you to the best medical center. You've chosen that, and I don't think that's a problem of justice if you deliberately put yourself beyond reach. It's a different matter if you're poor and you live in the back woods of West Virginia because you can't escape that poverty. But I think you start off with the assumption that so far as possible you want to treat people equally, assuming they have not by their behavior made it impossible for you to treat them equally by locating themselves in such a remote area you can't help them. But once you've made that assumption, thereafter the question is how do you try to provide people with the best possible access within the limits of your available resources.

**Baumrin:** You don't really believe that we have a duty to treat them equally?

**Callahan:** This is precisely why I'm prepared to use age. As somebody once said to me, "Ninety years old, 90 minutes old, they're all human beings, we're all equal." Yet it seems to me a terrible thing if a 90-minute-old person has to compete with a 90-year-old for resources. Somehow one would want to feel that the 90-year-old has already won a magnificent victory by getting to be 90. Faced with resource constraints, it would not be fair to go on pouring resources into the 90-year-old as if that person had the same status as the one who is 90 minutes old.

### Should Virtue Be Weighed?

**Baumrin:** But why don't we use virtue instead? That's actually used often—virtue in the sense of a 90-year-old, a delightful fellow, he was a delightful fellow even if he might not be such a delightful fellow anymore, compared to a drug addict.

**Callahan:** A 90-minute-old crack baby.

**Baumrin:** That's not bad. But take an AIDS baby. That's better because the prognosis for both the 90-year-old and the 90-minute-old might be roughly the same. Should I use enormous institutional resources to palliate the 90-minute-old baby who has AIDS, as opposed to using the same resources for the care of this 90-year-old? I think my gut reaction is, the 90-year-old is entitled. I'm not going to be able to do anything for this youngster. Why should I bother? That's a waste of resources. We have some social duty to the other person.

**Gorlin:** There's no social duty to that baby? If you can stabilize an individual with therapy for maybe six months, a year, two years?

**Baumrin:** If the prognosis is nondismal, I probably do have a social duty. But if the prognosis is dismal . . .

**Gorlin:** But we don't know. It's a disease without an answer yet.

**Baumrin:** I don't know what I'm doing. Am I creating a million-dollar social liability by saving this 90-minute-old baby now? Is that my *social* duty? I don't think so.

**Callahan:** We might agree. Let's assume that we don't want to provide futile treatment for any patient. If it were a 90-year-old and there's nothing we can do, we would say stop. But compare the 90-minute-old child who has a good prospect, which will require a large amount of resources to bring about, with a 90-year-old for whom percutaneous transluminal angioplasty will work. Should you treat the two in absolutely identical fashion in the face of limited resources?

**Baumrin:** I don't know exactly what type of metaphor to use. With the infant you're looking at a future of possibilities which may or may not come off, and you may think that you have a social debt to it, although it's not clear to me that there is one. With the 90-year-old there is clearly a debt. The person has served our community, has served us for a lifetime. Maybe this particular person is the crook Willie Sutton, but we owe something to the ordinary 90-year-old.

**Callahan:** Is it an unlimited debt?

**Baumrin:** Yes.

**Callahan:** Regardless of the burden on society? So therefore it's okay to tax all young people to increase what they pay into Medicare in order to help this 90-year-old, who's had a wonderful life, to have an indefinitely more wonderful life?

**Baumrin:** That's the price we pay.

**Callahan:** No, it's not "we" who pay, it's the young people who pay.

**Baumrin:** We are the young people.

### Quality of Life?

**Gorlin:** Let me get the two of you to address another value judgment. What about the quality of life of the individual? Do we have any rights to examine that and make decisions? You've got two 90-year-olds, to get back to Dr. Callahan's point, an age cutoff. One 90-year-old is thoroughly competent, in charge of his or her life, able to navigate, communicate; the other 90-year-old is incapable of either, requires continuing care. They both get the same acute myocardial infarction, and both of them go on to cardiac failure. One person was not functioning in terms of any relationship to society except to breathe and take in food and eliminate. How do you make that kind of call?

**Callahan:** I think that's enormously difficult. I suppose my instinct would be to say, If someone is severely demented, or even worse in a persistent vegetative state, then it would be reasonable to take that into account in comparison with another person of a similar age who had a better prospect. But it's a hard judgment to make precisely because I don't think we want to be in the business of saying that some people are inherently worth more than others just by virtue of their mental state. At what point do you begin crossing a dangerous line here? That's going to be our hardest problem in the years ahead, as a matter of fact, when we're dealing with more people in a sort of twilight area who can go on and on. I would want to give the benefit of the uncertainty to the one with the better prospect for the widest range of human function.

**Barbasch:** Who decides that?

**Callahan:** Let me make a distinction here. If it's a matter of resource constraints and we know you're going to have to make such distinctions, I would want to see the distinctions made at some institutional or societal level. I would hate to see distinctions made arbitrarily on the basis of any individual physician's way of thinking about these matters. I'd rather see it done with some consistency to insure some relative degree of fairness and objectivity here.

But let's assume you now have some general guidelines. It will fall on the physician to live within the guidelines, but probably to make some tough calls even so. So I suspect that as physicians, you're going to be stuck with the problem one way or the other.

**Barbasch:** So our patient comes to Dr. Gorlin for help. Dr. Gorlin will decide how chronic and how debilitated her state is and how much benefit she will derive from any of the acute treatments—that's what you're saying? Then he's going to tell her, "I've got the ability to correct the problem, but it is not worth it in your case." Isn't that the bottom line?

**Gorlin:** Let me put in a physician's comment. If the illness is acute, there's always the possibility the patient can be returned to their previous state, whatever it was. I think the real dilemma, as I see it, is, what happens after you've done that? You've gotten the patient past this acute myocardial infarction, helped her reach 90 years of age. Now what happens?

How aggressive do you get? The cardiologist may say, we should do cardiac surgery; or the suggestion may be, we should do an angioplasty. She's barely functioning. On the other hand, she is functioning, and she's 90—again putting the question back to the age limit. Do you expend a resource that now threatens to be in excess of $25,000? Maybe we as individual physicians don't ask that question. So what do we do?

**Barbasch:** Dr. Baumrin, who does decide whether to treat this patient with her chronic illness?

**Baumrin:** At the point when the patient was only 67 years old—a spring chicken as these things go—we're going to treat her, and we're going to treat her aggressively. The question of what we do when the patient has now tacked on another 23 years I agree is a social decision.

In fact, the way I predict things going is that the various associations and institutions that are going to have to grapple with this decision are going to redefine death, and that will probably come within the next ten years. By redefining death, you'll be able to call a lot of people dead that aren't now considered dead, and then institutions will be able to save money by saying, "That one's dead, that one's dead, that one's dead," and those unused resources will shift over to people who are still called alive.

But that's the future. What do we do right now?

**Barbasch:** Who makes that decision?

**Baumrin:** The decision is eventually going to be made, after discussion at a national level, by state legislatures. State legislatures will, for example, define death by upper cortical functioning. When upper cortical functioning ceases, a person is dead. Then, if that doesn't save enough money, they may turn to redefining death as periods of consciousness and conscious functioning that are measured by EEG over a period of time; a lot of other people will fail the test, and you get to call them dead. This is the way things will go because, if everybody keeps saying "I can keep them alive by pumping their lungs, pumping their hearts" when they're really not alive—and that's what people have been saying now for ten years—legislatures will get the message that we're keeping dead people alive or going, at great cost to society, so let's stop it. Forces are pushing for that now, and those forces, I think, will ultimately be successful, because there's big money associated with it.

When it comes to what to do with the patient now, Gorlin is . . . are you still treating her 23 years later?

**Gorlin:** Of course.

**Baumrin:** Gorlin is treating this patient, and she needs TPA, and it costs $25,000 a year; he thinks nothing of it. He's going to let the senior hospital administrator who has to take care of these things worry about where the $25,000 is going to come from, because he's not an accountant, he's not an administrator, he's a physician. If this is what she needs, this is what she gets. Now, if people around want to stop it, they have to say, You can't do this anymore; they have to stop him, he does not have to stop himself.

**Barbasch:** How about Callahan stopping it?

**Baumrin:** I don't know. Is he functioning as a physician, or as the chief hospital administrator, or the state's primary ethicist? In the latter two cases he may stop it.

### Regulations by Cost?

**Callahan:** What I would do if I were in a position to do it and have to set the regulations would be first to call in Dr. Gorlin and say, "Dr. Gorlin, I've got to set some rules about these matters. You've got to tell me clinically everything you know about treating elderly patients so that I can do this in the most reasonable way possible." I wouldn't want to do it arbitrarily without consulting the medical community that has to live with the result. So that, we hope, any decision would be one that physicians think, given the resources and restraints, is a reasonable kind of decision. It would be one that has taken account of a great deal of clinical experience. It would be painful to apply but nonetheless we've done everything possible to make it as least painful as possible.

**Gorlin:** You see I think also the phrasing "least painful as possible" applies to the patient, because if that 90-year-old patient is in subjective discomfort, severe and unremitting discomfort, and I cannot use a so-called cheap treatment to effect recovery, then I think I'm obligated, by being a physician treating a suffering patient, to go to an aggressive therapy if that's what will relieve the pain.

I think the dilemma comes up when the patient is fine but she's going into heart failure, and I've been told by my consultant that we might be able to make her heart failure less, and maybe allow her to live a little longer, if we did a bypass procedure or an angioplasty. But she's fine, she's comfortable. That's where I hear clearly your suggestion that, if you can keep her comfortable, do you have to do every single aggressive thing? I think it's a challenging philosophy. If I can generally keep the patient comfortable, I can get her home, get her on diuretics, get her a home health aide and all that society is willing to see that she gets, she's not cast out on the street. I can live with that as a physician.

**Callahan:** What about the first case? What would be your advice to me as a regulator about this first case—That we should simply, given the resource restraints, swallow the costs, even if they are enormous?

**Gorlin:** I think it depends on whether the gain is large or small. If the gain is large, I think you have to address the cost the same way Medicare has addressed transplantation when it was found to be saving lives under certain conditions, and the costs are enormous; I think you have to do the same with any other treatment. If the gains are at the margin, then I think society has a tough job of arguing that through. In any case, the argument should not be taken lightly.

### Questions and Answers

**Barbasch:** Our time, unfortunately, is up, and we will now entertain questions from the audience. Dr. Levitt.

### What Care for Antisocial Patients?

**Marvin Levitt:** You discussed the 90-year-old woman and the 90-minute-old child, but we have a great problem with another group of patients: patients who are documented psychopaths with evil, destructive histories. I'll give you two examples.

The first occurred at the Veterans Administration. I was watching a dialysis patient have a second aortic valve put in. The police surrounded him because of his history of murder and rape. When the operation was completed, he was going back to jail.

The second was a dialysis patient who used his vascular access to facilitate self-administration of intravenous drugs; he was destructive and disruptive to the nurses and was not compliant. He often came in with potassium levels in excess of 7.5 mmol/L. Yet we were obligated to treat him.

What is your attitude toward treating such dysfunctional patients and using our facilities for people with documented evil, asocial, psychopathic personalities?

**Barbasch:** Dr. Baumrin, I'm sure that's directed to you.

**Baumrin:** I actually think the virtue test I mentioned isn't totally silly. There is for antisocial individuals a point at which one says, "We've done all we need to do for you. We don't owe you any more." I don't know exactly how you make the judgment, but you gave two good cases of where the judgment wouldn't be too difficult. I don't think that you are endlessly obligated to treat people who are self-destructive, particularly self-destructive with illegal material like drugs. That doesn't mean that you don't make, say, palliative care available, or minor, relatively costless care. But there's no point in doing anything big for them.

Nevertheless, the judgment that they're not entitled is not a judgment that you as a physician ought to be making. The judgment that they're not entitled is a decision the society should be making, so that society can face itself on what it's doing. Now, the social decision has been made to treat everybody equally, obligating everybody who engages in medical care to treat patients as if they were all terrific, just wonderful people. The only thing you can do, the most you can do individually, is pass them on—to Gorlin, to Barbasch—pass them on because you don't want to deal with them. But our society refuses to face this question at the level that Callahan is suggesting we face the other issues, which is at the legislative level. It wrenches me to think of my 90-year-old having his care curtailed while endless resources are served up to sociopaths, and to what end? So that they can injure others and themselves yet again? Why do they deserve it?

**Callahan:** There's a fundamental moral issue here. The tradition of medical ethics has said that you should treat all patients who come before you without judging why they got there or what kind of people they are. The large issue that's on our mind right now is the likelihood of a war with Iraq. We would expect the American physicians, if they come across Iraqi wounded, to take care of those wounded and to give them the best treatment possible, to treat them as individual patients, and forget the fact that a minute before they were trying to kill Americans. That's a profound part of the ethic of medicine, difficult to start compromising or to know how to compromise.

You as a physician cannot stand at the door and try to pass judgment on people's moral character. You may make many errors, even about people who seem behaviorally the worst. They may be psychotic, they may be driven psychologically in ways beyond their control—it's tricky business. We might, though, want to set some tough standards for people who repeatedly put themselves in harm's way, those with a track record of repeated harmful behavior—although I think that's hard to do also. But I don't think we can judge character. You would really have to overthrow a fundamental tenet of medical ethics to do that, and that would be a dangerous thing to do; it would lead to enormous injustice. We also have different notions of what counts as evil in this world. When I show up at the emergency room I don't want to have my physician say, I hate those people who want to set age limits on the old.

**Barbasch:** While we have tried to deal with some of the dilemmas of daily practice, we have raised many more issues. As our knowledge increases, we will be forced to deal with more complex problems. We must remain on guard that we not adopt simplistic solutions with terrible consequences as we grapple with these problems. The measure of our greatness will be determined by how we treat the weakest among us.

My thanks to the panelists for their enthusiastic participation.

# The Personal Ethics of the Physician:

## Curing Medicine from Within

EDMUND D. PELLEGRINO, M.D.

COMMENCEMENT speakers have two obligations—one pleasant, one much more dubious. The pleasant task is to congratulate the new graduates and wish them well. This I am happy and honored to do. I must also congratulate your families, faculty, and friends because in their way they too have contributed to your accomplishment. You should all be proud of your shared efforts in one of humankind's nobler accomplishments—the making of a new doctor.

Despite the gloomy portents all around us, I want to welcome the new graduates enthusiastically into what I still believe to be a high calling, and a most satisfying and useful way of life. Forty-five years in medicine have not blunted my enthusiasm for our profession or its practitioners. With all our faults, we are still among the few groups in our society committed to some degree of effacement of self-interest in the interest of our patients.

This assertion takes me directly to the second of my tasks, the less welcome one, the commencement address, that peculiar obsession of academia which dictates that the joy of getting your degree be touched with a little pain. I must administer the last penance your medical school requires of you. The only antidote is brevity, and the merciful amnesia that will quickly erase the speaker's name and topic even before you leave this ceremony. There is, you see, no danger of permanent brain damage.

In a few minutes you will receive your degrees. This is the University's public declaration that you possess special knowledge. You should be proud of that degree. But it is not sufficient to make you a doctor. You will really enter the profession when you take the Hippocratic Oath. This is your public declaration, your promise that you will use your knowledge for the sick and not primarily for your own interests. This is what sets you apart as a true professional.

Before you take this solemn oath, let me ask you to reflect on its significance for your life as a doctor. To take an oath is a solemn act. In the moments remaining before you do so, I ask you to reflect on what the oath symbolizes and why it has for 2,500 years called doctors to a higher set of obligations than we expect of most people in our society.

If you do not take the oath with sincerity today, the rest of your medical life will be a lie. If you do, you will have to give an account every day of your fidelity to its ethical imperatives. Every day of your professional life, your human inclination to power, pleasure, prestige, and profit will tempt you to place self-interest ahead of your primary ethical obligations. It is easy to lose your way when so many around you are making moral compromises.

### Six Moral Compass Points

Let me suggest six unchanging realities about being a physician that will be your compass points—the reminders that will set you back on track when you begin to lose your way, when you are confused about what you are—healer, technician, gatekeeper, bureaucrat, corporate employee, and a dozen other roles being pressed on us. It is my conviction that much of the moral malaise, the loss of enthusiasm and self-worth I see in my colleagues is the direct result of a failure to take account of these compass points.

---

Adapted from the author's Commencement Address at the Mount Sinai School of Medicine (CUNY) graduation ceremony on May 18, 1990 in New York City. The author is Director of the Center for the Advanced Study of Ethics, Georgetown University, Washington, D.C., 20057-1084. Address requests for reprints to the author there.

**The Vulnerability of the Patient.** The first compass point is the dependence, vulnerability, and exploitability of the patient who seeks your help. A sick person is anxious, in distress, and fearful. To be healed or helped, she is impelled to seek your help though she wishes she could avoid it. The patient is not free to pursue life's other goals until health is restored or symptoms alleviated. Will you forget the uniqueness and the vulnerability built into the human predicament of illness?

**Inequality of Power.** The second compass point is the inherent inequality of the relationship. You possess the knowledge the patient needs. The preponderance of power is in your hands. You can use it well or poorly, for good or evil, for service or self-interest. You have enormous power to modify, modulate, and manipulate the patient's decisions. Will you use that power to impose your values, your social or political philosophy?

**The Covenant of Trust.** The third point about medical relationships is the special fiduciary character of the healing relationship. In a state of vulnerability and inequality, patients are forced to trust physicians. Patients are ill equipped to evaluate our competence. They are forced to reveal their intimate selves—baring their bodies, their personal lives, their souls, and their failings to another person who is a stranger. They trust us with these invasions of their privacy, because without trust they cannot be healed or helped. Moreover, we invite their trust. We begin our relationship with patients with an invitation. We ask, "How can I help you?" Implicitly we are making a promise: "I have the knowledge you need. Trust me to have it, and to use it in your best interests." Will you be faithful to this trust, which is central to all codes of ethical medicine, in all times and cultures?

**Our Nonproprietary Knowledge.** Fourth, the physician's knowledge cannot be wholly proprietary. It is knowledge ordained to a practical end, to meeting certain fundamental human needs. Medical knowledge does not exist for its own sake. Moreover, society permits us to gain that knowledge through invasions of privacy that would otherwise be criminal. In order that physicians may be trained, medical students are permitted to dissect human bodies, attend and assist at autopsies and operations, and participate in the care of sick people. Society permits these privileges not primarily so we physicians can make a living, but because society needs an uninterrupted supply of skilled doctors and because there is no substitute for hands-on medical training. From the first day of medical school when you accepted the privilege of a medical education, like it or not you entered a covenant with society. Will you hold your medical knowledge in trust for those who need it? Or will you dispense it primarily for your own profit?

**The Role of Moral Accomplice.** The fifth unchanging reality of the medical relationship is that the physician is the final common pathway through which help and harm must pass. Nothing happens to our patient without our orders. No policy, no law, no regulation can be effective without our cooperation. We are the final guardians of the patient's interest and de facto moral accomplices in what happens to the patient. We write the orders no matter what the policymakers or the law may require. Will you do the Pontius Pilate act and lay the blame elsewhere?

**Membership in a Moral Community.** The sixth distinguishing characteristic of medical relationships is that the doctor is a member of a moral community; collectively we share the privileges and obligations of our special knowledge. We are granted a monopoly on medical knowledge and hold it collectively as well as individually in trust for those who need it. We are collectively responsible for its equitable accessibility, availability, and distribution. We are responsible to and for each other. Will you, as a physician, be a moral island sufficient unto yourself, unmindful of your responsibility as member of a moral community?

These are your six compass points (vulnerability, inequality of power, the covenant of trust, the nonproprietary nature of our knowledge, the role of moral accomplice, and membership in a moral community). They are the ineradicable facts that generate the obligations the oath makes public. These are the ineradicable components of the "internal morality" of our profession, the immediate moral ground for all our obligations. They are the sources of the virtues required of a good physician.

## The Role of Personal Ethics

In the current rebirth of medical ethics we have overemphasized dilemmas, quandaries, and puzzles. We have given the impression that ethics consists only of the right decision in a complicated case. We have ignored the personal ethics of the physician and forgotten that ultimately everything depends on the physician's virtue and character. We do need healers who know how to analyze moral dilemmas. But it takes more than philosophy or ethics to dispose us to do what is right and good for a sick person.

Sometimes in your own professional lives each of you will enter into a Dark Wood you will have to face yourselves. You will have to answer for where you are, what you have, and who you have become. Will you be able to say you have been a faithful steward, or will you have betrayed your trust? I hope the answer does not come only when you too are ill and face death, dependence, vulnerability, exploitability, and the need for healing.

Our profession is in a quandary today. We point to malpractice, government regulation, competition, commercialization, advertising, our personal debts, AIDS, for-profit medicine, and blame them for our malaise, our loss of enthusiasm. Some even feel justified in striking back, refusing to care for certain patients, "making it" while we can. But our illness is an illness from within. And it is curable only from within; only by recapturing our moral integrity can we recapture our identity and the joy that comes from a life spent not in pursuit of self-interest, but in service to the sick.

That is why the most important prescription of all is the ancient one we write for ourselves—"Physician, heal thyself." The six compass points I have suggested are the ingredients of that prescription and that healing. Keep them in mind as you take the Oath today, ponder them every day, and you will lead a life of satisfaction, service, and moral integrity.

# Medical Ethics: A Broader Definition

Sidney M. Wolfe, M.D.

Filling the shoes of Dr. David Axelrod, New York State Health Commissioner since 1978, who was to have given this lecture but whose illness forced cancellation, is impossible, either as a speaker on medical ethics or as a health commissioner practicing medical ethics as no public official has before in this country.

I asked three people—a doctor, a lawyer, and a chief—who have worked with Dr. Axelrod, What does Dr. Axelrod mean to you now that you realize that he is not going to be around as health commissioner anymore? The lawyer said, "He is an enormous loss—he had vision, he had courage. I was privileged to work for him. He so clearly set a tone, that we were public servants, an elevated calling, at a time when 'big government' is being so bashed."

A physician who knew Axelrod well, worked with him closely, and whose career was clearly shaped by him said, "Not only was he brilliant, but he personified public service and what a public servant should be"—the same theme, that government service is a noble cause.

The chief, who has worked under Dr. Axelrod for many years, said, "I observed that for him medicine has been the highest calling; his personal love and dedication for the profession has been the driving force behind his actions to increase the quality of medical care. I shall miss his willingness to stand firm in the face of opposition when he believed his course was serving the public interest. He was prepared to risk his career, rather than yield to self-serving groups—such as the New York State Medical Society—who had significant power. Dr. Axelrod has always demanded the best of himself and everyone who served with him. He set the standard in the department for integrity, dedication, professional discipline, imaginative resolution of pressing public problems, responsiveness to individuals, and public accountability. His extraordinary ability to grapple with global public health issues is unparalleled in my experience. Yet he could turn in an instant to delve into the personal needs and problems of a clerk or technician. To me, he has embodied all the highest attributes of the calling of medicine. He has been New York's personal physician."

As one would expect of a doctor on call to everyone in New York State, he had a listed phone number, as an April 1991 newspaper editorial (1) noted.

Some of the exemplifications of serious ethical problems in medicine that I discuss here are personal ethical problems that physicians individually can act on to bring about change; some are more global and societal and are foisted on society by, for example, the drug and medical device industries. Physicians should at the least be interested in these more global problems and should try to act on them to bring about change. My theme throughout this discussion is that we physicians have a dual responsibility—not only to practice and do research ethically, but also to spend some of our time on broader societal change.

In looking at the quality of the health care system, just as I am using the word "ethics" in a broader sense than is traditional, I also use the word "quality" in a broader sense. Problems arise unless the health care system is attending to quality in prevention, in access, in diagnosis and treatment (the most usual focus for discussions of quality), and in quality review. Thus, quality

---

Adapted from the author's Alexander Richman Commemorative Lecture in Humanism and Ethics in Medicine at the Mount Sinai School of Medicine (CUNY) on May 1, 1991. From the Health Research Group, Public Citizen, Washington, D.C. Address reprint requests to the author, Director, Health Research Group, Public Citizen, 2000 P Street, NW, Washington, D.C. 20036.

must be maintained in all four respects. I present here examples in each of these four areas, probably more in the area of quality of diagnosis and treatment because the Health Research Group does more work in this area. I argue, however, that we as physicians, and the health care system in its entirety, are seriously failing in making our health care system fulfill any of these four aspects of quality.

**Quality in Prevention.** Any campaign to improve prevention has several elements: occupational exposures, environmental exposures, medical exposures (iatrogenic), lifestyle-induced disease, and immunization and screening. If we do not prevent disease and therefore wind up treating it, or people die because we did not prevent the disease, we are not doing a good job.

A single example is the connection of aspirin and Reye's syndrome. Unfortunately, that connection has resulted in thousands of dead children and many thousands more of brain-damaged children, because of the greed of the aspirin industry, people in government, and others whom industry was able to convince to go along with them. Aspirin, when taken in conjunction with chicken pox or flulike illness, greatly increases (by twentyfold or thirtyfold) the odds that the taker will get Reye's syndrome. The odds are so high that aspirin should not be used, and now mainly is not, for treating chicken pox or flu. Reye's syndrome is therefore a preventable disease.

The chronology of events is revealing. In October 1981, a Centers for Disease Control committee advised against the use of salicylates—aspirin—for treating chicken pox or flu. By February 1982, having reviewed the case control studies on which the Centers for Disease Control advice was based, the Health Research Group, which I have directed for almost twenty years, decided that the evidence was more than sufficient for the government to follow the drug safety laws and require a warning label on all products containing aspirin. The Health Research Group petitioned the Food and Drug Administration (FDA), incorporating a review of all published and unpublished studies, and stating our belief that, unless the government wanted to break the law, a label on every aspirin bottle had to state that it should not be used by children or teenagers to treat chicken pox or flu. Because the FDA did not respond, we sued that agency. In September 1982 it seemed to be responding, proposing a mandatory warning label on all aspirin-containing products. The drug industry, the aspirin industry particularly, did not like this stance and claimed that the studies really were not perfect—How could you do something when the data are not perfect? The White House and the Office of Management and Budget killed the labeling proposal in November 1982. We continued our lawsuit, and we also encouraged members of Congress to introduce legislation requiring a warning label. After a large number of children had been killed, in March 1986 the Food and Drug Administration, having failed at persuading the industry to voluntarily put a warning label on the product, required the drug companies to label all aspirin-containing products.

Between 1977 and 1980, the average number each year of reported cases of Reye's syndrome (the actual number of cases was much higher) was 391; 20%–40% of Reye's syndrome cases are fatal, and many other patients sustain brain damage. In 1982, because we and others were loudly trying to educate the public and parents, despite the absence of a warning label, the number of new cases diminished. I must have spent 60%–70% of my time for a year or so working on this issue. By 1984, because no labels were yet in use, the number was not much lower. Finally, in 1986, the year the warning labels became mandatory, the number dropped to 101; in 1988 the number was 20; and in the last year the number of cases dropped to 12. Reye's syndrome is an almost entirely preventable disease. Delaying the warning labels was completely unethical. The companies have had a small number of lawsuits compared with the amount of damage, but that is really after the fact. The reason the number of suits is not larger is that few parents who gave their small child aspirin when the child had chicken pox or flu, and thereby contributed to their death or mental deterioration, are willing to involve themselves in an unpleasant legal tangle as a result.

That is one of many examples in which prevention should have been a primary defense. Smoking is another glaring example. Many universities and their medical schools continue to invest heavily in the tobacco industry. The profession of medicine has not done a good job in helping the 51 million smokers who are addicted to cigarettes to stop, even though some methods seem to work better than others.

**Quality in Access.** This country is doing a rotten job in improving access to health care, whether measured as access to the system through insurance or other coverage, or access within the system (can the patient connect?), or even in measurement of barriers to access. Two recent papers, one already published (2) and the other not yet submitted for publication (3), examine the problem of why a country like the United States will spend $756 billion on health in 1991, yet have 35 million people who remain uninsured. One of the answers is that we waste an extraordinary amount of money on oppressive administrative costs—oppressing doctors, who, compared with their Canadian counterparts, spend much more money and time filling out bills for the 1500 different U.S. health insurance companies. Our health care system is also oppressive at the level of the hospital; 10% or more of the budget for many hospitals constitutes administrative waste. So 14% of the people in this country do not have health insurance (2), yet 15%–16% of our health care expenditure is on administrative costs that are not wasted in the more efficient Canadian system. We need to do something about that. The United States stands with South Africa as one of the only two industrialized countries that do not really believe that being a citizen gives you the right to health care.

**Quality in Diagnosis and Treatment.** Quality in diagnosis and treatment can be measured by answers to these questions: (a) Is the history and physical examination well done? (b) Are tests appropriate? (c) Are hospitalizations and operations appropriate? (d) Are drug prescriptions appropriate? (e) Is quality of interpersonal care high?

One example of a serious ethical problem that affects the quality of drug prescriptions is an advertisement by a major U.S. pharmaceutical company of the sort that slips through the cracks because these ads are not cleared by the FDA, which does not have the resources to do so. The ad shows an air traffic controller sitting at a computer terminal looking worried (as one would expect, because 10,000 of his colleagues were fired a few years ago). It seems to me the remedy for his anxiety would be to hire more air traffic controllers. But the company producing the ad wants to push tranquilizers to treat the anxiety of air traffic controllers. (I received this ad in the mail. I am an internist and, much to the chagrin of the pharmaceutical companies, I am on all the computerized mailing lists that all of you are on, so I keep getting ads like this in the mail.) This ad essentially told me the tranquilizer Buspirone is so safe that you can prescribe it for an air traffic controller; granted it is less sedating than benzodiazepines, but it is hardly the proper choice for an air traffic controller, as the small print on the third page of the ad, which those of us over 50 have to squint or use a magnifying glass to see, confirms. It warns that the drug should not be used by people who are driving automobiles or operating delicate machinery.

We at the Health Research Group were able to get the ad stopped in a day or so, but what would have happened if one air traffic controller, prescribed this drug by a physician, had had some central nervous system or visual problem, a blackout, or whatever, and a plane had crashed and hundreds of people had been killed? The company that put the ad out would have been, could have been, and should have been eliminated from the face of the earth through liability suits, but that is after the fact also. Why do companies run ads like that? Because the drug was not selling well and they figured that by making it appear safe enough for air traffic controllers, it would sell better. This is hardly the only example of unethical or illegal drug advertising and, in my view, exemplifies the criminal and unethical way in which drug companies try to sell their products.

What do these people have in common: a 58-year-old man with parkinsonism, a 68-year-old woman with a hip fracture, a 63-year-old woman whose memory and thinking are failing, a 52-year-old woman who suddenly dies during oral surgery, and a 63-year-old man in a car crash? Each is a well-documented, specific instance of an adverse drug reaction in a more-or-less older person. The 52-year-old I don't want to call "older," because that is younger than I am, but the other ones are at least beginning to be "older" people.

Are these isolated instances? I do not think so. Studies in this country on the magnitude of adverse drug reactions in older adults (4) add up to 243,000 drug-induced hospitalizations; 32,000 hip fractures occasioned by falls caused by the use of tranquilizers, sleeping pills, or other drugs; 163,000 people with drug-induced or worsened memory loss or dementia; 73,000 people with tardive dyskinesia; and 61,000 people with drug-induced parkinsonism.

We think two thirds of the prescriptions for older adults are wrong for a variety of reasons: (a) No prescription is needed (mild hypertension; adult-onset diabetes, for which the first resort should be weight loss or reduced caloric intake). (b) Unnecessarily dangerous drugs are prescribed; *Worst Pills, Best Pills* (4) notes that of the 287 drugs most commonly prescribed for older adults, 104 drugs should be considered "do not use" agents for elderly persons. (c) The right drug is prescribed, but the dose is too high—probably the most common prescribing mistake, because the FDA has not pushed the industry to educate physicians that for almost all drugs, risk can be reduced if the rule "start low, go slow" is followed; dosage should be increased only if necessary.

One of Dr. Axelrod's campaigns was directed at the triplicate prescription program for controlled substances, in this case benzodiazepines. According to New York State Health Department data, a small decrease of 9.9% occurred nationwide in the number of prescriptions for benzodiazepines filled in retail pharmacies from 1988 to 1989 (1988, 82 million prescriptions; 1989, 74 million). But in New York State, within one year after the triplicate prescription program began for just the Medicaid program, the decrease in prescriptions was 59.3% (1987, 1.69 million prescriptions; 1988, 1.56 million; 1989, 0.63 million). Some but not much of this decrease was offset by a switch to other drugs—chloral hydrate, for example—but by and large most benzodiazepine prescribing disappeared. Most of the change was scattered throughout the state and reflects, I think, more caution by doctors, who, as they press a little harder to get triplicate prescriptions, realize that maybe it isn't a good idea to contribute to the 3 million people in the United States getting daily prescriptions for tranquilizers or sleeping pills for a year or more, even though there is no evidence that the sleeping pills work for more than 30 days or the tranquilizers for more than 4 months. I hope that this remarkable program can be expanded. Some glitches seem to have been worked out, including the issue of refills.

Dr. Axelrod has also pioneered in the collection and dissemination of data on what is going on in medicine, for example data from New York State Health Department for 1989 on doctor-specific bypass surgery death rates. The expected death rates, which take account of how severely ill the patient is (as agreed on by a number of surgeon specialists who worked with the state), have a relatively narrow range, from 2.73% to 6.23%. But the actual death rates ranged as high as 18.18%. The surgeon with that highest percentage of actual deaths posted an even higher mortality rate of 22.3% when the data were adjusted for severity of disease.

This is an ethical issue, just as not knowing enough about prescribing drugs and thereby causing people to fall or get parkinsonism is an ethical issue for the physician. Here, however, the ethical issue is at the level of the physician and of the hospital. Assume that you have severe coronary artery disease and have been properly studied diagnostically and that it has been determined, after a second opinion, that you need coronary artery bypass surgery. Even though some of those operations are unnecessary, assume this operation is necessary. Someone tells you that a dozen or more hospitals in New York City can perform the procedure. In one hospital, one of the surgeons performs a large volume of such procedures, and 1.61% of those who have the operation performed by that surgeon die, a lower rate than could be expected given the severity of illness of that surgeon's patients. In another hospital is a surgeon who doesn't perform as many operations annually—55 procedures a year—whose procedures have an actual death rate of 18.18%.

Which hospital would you choose? "The low one," of course. But until recently your attempts to get the surgeon's and the hospital's name were futile—"Sorry, we can't tell you which is which—that is secret, that invades the privacy of the doctor or the hospital." Dr. Axelrod and others have pioneered the idea that the public has a right to this data, particularly if the data are valid.

The reasons why we think unnecessary cesarean sections constitute a major ethical dilemma in American medicine, and why we devote an enormous amount of time to the issue, are these:

- the most common reason for hospitalization in the United States is to deliver a child—4 million deliveries a year.
- The number one major surgical procedure is a cesarean section—almost 1 million a year. The rate is now 24%; in the early 1960's when I was in medical school it was 5% or 6%. The rate probably should have risen to 10% or 12%, not to 24% or higher.
- The most common unnecessary operation in the United States is cesarean section (5). We estimate that about half of these 475,000 procedures are unnecessary. This procedure wastes $1–2 billion and causes many serious infections, as any major abdominal procedure would do. It is particularly unfortunate since many of these women could have had vaginal delivery.

In our opinion, the U.S. hospitals that have the highest ratios of cesarean sections to deliveries are Hialeah General, FL, 53.1% (of 2239); Montgomery City, IA, 50.0% (of 56); Palm Beach Garden Medical, FL, 47.7% (of 983); Kent and Queen Anne's, MD, 47.0% (of 611); Allen Bennett Memorial, SC, 46.5% (of 245); South Miami Hosp. Foundation, FL, 46.4% (of 3055); McCready Memorial, MD, 45.5% (of 53); Holy Cross, FL, 45.2% (of

1598); Miners, PA, 44.8% (of 67); Coffee Med. Ctr, TN, 44.7% (of 76).

The hospital that heads the list with a rate of 53.1%, Hialeah General Hospital, did not have a neonatal intensive care unit; they were not taking care of a particularly ill set of women who had not had prenatal care. North Central Bronx (NY) Hospital, unlike Hialeah General Hospital, has many patients who have not had prenatal care and are at high risk; however, their cesarean section rate is about 9%.

An exciting recent paper (6) describes a new way of lowering the cesarean section rate. If a doula—a birth assistant who helps a woman through labor and delivery with explanations and hands-on caring—sits and watches and explains to and touches the woman, as in hospitals that use many nurse-midwives, the cesarean section rate is much lower.

The main reasons for unnecessary cesarean sections are financial incentives for hospitals—a third of the hospital beds in this country are empty, and a 5-day admission lasts longer than a 2½-day admission. The hospital bill is $2–3,000 higher per case for a cesarean section than for a vaginal delivery. Financial incentives for doctors, much smaller than for hospitals, also contribute. Convenience for hospitals and for doctors is a factor. Particularly in small hospitals, but even in large ones, when it is midnight and no one else is around, a cesarean section is the choice; a nurse midwife or a doula could alter this situation. Automatic schedulings of repeat cesarean sections also contribute. Perhaps 70%–80% of women who have had one cesarean section could have a vaginal delivery the second time. However, only about 11% do. Doctors automatically schedule a repeat cesarean section on finding out that the woman is pregnant. Unsafe reliance on fetal monitors, impatience of doctors, ignorance, or bad habits are other reasons. A resident training in obstetrics and gynecology at Hialeah General Hospital will think that it is OK to do cesarean sections 53% of the time.

I do not think malpractice fears are relevant, although clearly physicians think they are, and I do not dispute their beliefs. But several lines of reasoning, of evidence, make one wonder how important a factor malpractice is. One is that a significant number of unnecessary cesarean sections are in women who never go into labor; they are automatically scheduled for repeat cesarean sections when their pregnancy is known. They are hardly at high risk; they should not be treated differently from women who have never had a cesarean section. Second, a study in California shows a significantly higher cesarean section rate in for-profit hospitals than in non-profit hospitals. The doctors in all those hospitals pay the same malpractice premiums.

Among cures for the cesarean-section epidemic are equalizing reimbursement for uncomplicated cesareans and vaginal deliveries; instituting hospital protocols that limit cesarean section to only the necessary ones (every hospital with such a protocol has cut the rate in half, down to the 10%–12% rate that we and others believe is appropriate); allowing women who have had a cesarean section to have a trial of labor. Other cures include improving informed consent, and educating doctors about the malpractice risks of unnecessary cesarean sections. A small but growing body of litigation, again unfortunately after the fact, has been built up in suits brought by women who had an unnecessary cesarean section and sustained a major complication as a result.

**Quality in Quality Review.** Quality review depends on high quality in credentialing new applicants, in reviewing performance, in disciplining and probation options, in JCAHO operations, in professional review, and in the work of state licensing boards and state health departments.

Let us examine data on just one of these factors, state medical boards' disciplinary actions on doctors. The rates of serious disciplinary action by state, in number of actions (revocations, suspensions, probations) per thousand doctors, in 1989 revealed a wide disparity between the states with the highest (left) and lowest (right) rates:

| | |
|---|---|
| Missouri, 7.02 | Idaho, 1.45 |
| Georgia, 6.85 | Virginia, 1.35 |
| Mississippi, 6.63 | Vermont, 1.26 |
| Oklahoma, 6.12 | Maine, 1.21 |
| West Virginia, 5.89 | Washington, 1.01 |
| Iowa, 5.43 | Maryland, 0.98 |
| Alaska, 5.3 | Texas, 0.94 |
| Colorado, 4.98 | North Dakota, 0.84 |
| South Carolina, 4.87 | Connecticut, 0.48 |
| Indiana, 4.84 | New Hampshire, 0 |

Can the doctors in New Hampshire, for example, be better by 15-fold than those in Missouri? We have no evidence for that, or that physicians settle geographically as a function of their competence. What we do have evidence for is that some states take the process of disciplining doctors much more seriously, as what we believe is an ethical obligation, than others. Some fund their licensing and disciplinary boards better, have more investigators, have more subpoena power, and in sum do a better job.

But this is hardly the only failing component of quality review. The Joint Commission on Hospital Accreditation accredited huge numbers of hospitals that failed to pass hospital inspection by New York State; the same discrepancy occurs in some other states.

**Doctor-Bribing and Quality in Treatment.** The Health Research Group has spent much time on documenting and curbing doctor-bribing, bribing in the criminal sense of getting doctors to do something for money or some other perquisite. In 1990 I started a hot line for physicians and others to report to me instances of such efforts. The following memo, verbatim, was sent out by the administration of a hospital in suburban Chicago a little over a year ago:

> To: all staff physicians and surgeons
> From: the vice president of the surgical service
> Any physician who performs ten major surgery cases between November 23, 1989 and January 31, 1990 will be entitled to receive by March 31, 1990 their choice of one of the following promotional incentives: an in office fax machine or a portable cellular telephone.

This attempt to bribe doctors to fill empty beds in the slow season for elective surgery is criminal activity, which was stopped before it resulted in any patient being operated on. It violates a 1977 antikickback law, which affects Medicare and Medicaid patients.

Is it an isolated instance? No. We have documented several other examples. One, on which I have all the documents, was a frequent flyer-frequent prescriber program. Long-acting Inderal, a version of a good drug, propanolol, wasn't selling well. The Wyeth-Ayerst division of American Home Products worked out a deal with American Airlines that any time a doctor prescribed long-acting Inderal, that doctor got 1000 frequent-flyer bonus miles on American Airlines. Fifty prescriptions brought the doctor a free coach ticket to anywhere in the United States. American Airlines refused to tell how many doctors participated but said the program was very successful. This is criminal activity to the extent that it involves Medicare and Medicaid patients. The company has already coughed up almost $200,000 to the Massachusetts Attorney General, who prosecuted the case. They are currently under active investigation by the Inspector General of the United States, Department of Health and Human Services.

A second example involves cash bribes for prescribing Rocephin. The deal is that the doctor takes a minute or two to fill out a piece of paper that says how old the patient was and what their infection was, and gets $60 cash. The doctor who does the same thing twenty times gets $1200. Not a bad deal, because the 20 prescriptions that bring the doctor $1200 cash yields $10,000 in sales of Rocephin, an expensive antibiotic. This is illegal criminal activity that is under investigation.

How about a $30,000 computer system for your office to take care of all of your patient records and to monitor your drug-prescribing practices, with only a couple of hitches: one, you have to watch a prescribed number of drug ads every month; two, the company can watch your prescribing practices to see whether their drug ads are working. This also is a successful program.

The most widespread of these practices, what our Health Research Group has called "wining, dining, and pocket lining," involves most major pharmaceutical companies and about a third of all the doctors in this country. Drug companies don't want to interrupt the doctor, because the doctor has to see patients. The detail person interrupts the doctor. The pharmaceutical company reasons, Let's take the doctor away from the office for an evening symposium. The offer used to be just wining and dining, which alone was questionable, but increasingly now there is pocket lining, as in $100, $150, $200 cash to induce the doctor to go. Does it work? You bet. The companies putting on these symposiums say that, in 1989, 170,000 doctors went to one or more; 80% of these doctors increased their prescriptions for the drug discussed at the symposium.

Traditional discussions of medical ethics tend to follow lines of reasoning on death and dying and a number of other serious issues. At the Health Research Group we are concerned about these issues too. But it is important to broaden, as I have done here, the topic of medical ethics to include other issues clearly in the arena, yet unfortunately not always viewed as questions of medical ethics. An enormous amount of denial goes on, both by pharmaceutical companies, who try to bribe doctors and sometimes succeed, and by doctors, who take free trips and accept other perks. Denial is a powerful mechanism of defending the ego against having to believe that you are doing something wrong.

**Summary.** The definition of medical ethics needs to be expanded. But in addition, medical ethics for all of us in medicine, which includes physicians, researchers, nurses, nurse-practitioners, and anyone who participates in health care, always extends to two levels.

One, we have to be ethical in our own practice and in our own research. We have to make sure that whatever we are doing in diagnosis and treatment of patients is not influenced by anything but what is in the best interests of that patient—not by wining, dining, and pocket-lining, or any of the other vulgar activities—disgusting is the only appropriate word—increasingly going on in medicine. Being ethical at a personal level will work for those people who accept this ethical responsibility and carry it out.

Two, we also have to spend some time, not necessarily a great deal, being ethical at a societal level—time worrying about living in a country that has a health-care system that denies so much care, worrying about our professional societies not really coming down as hard as they should on those members not practicing good medicine or doing other things that are unethical, if not illegal or dangerous.

If we in medicine fail to accept both of these levels of ethical responsibility, our health care system, at the best wonderful, but at the worst terrible, will continue to deteriorate. If we really work, though, at changing those faults, and constantly focus on both levels of ethical behavior, our system will get better.

### References

1. Albany Times Union, April 19, 1991.
2. Woolhandler S, Himmelstein DU. The deteriorating administrative efficiency of the US health care system. N Engl J Med 1991; 324:1253–1258.
3. Hellander I, Himmelstein DU, Wolfe SM, Woolhandler S. Administrative waste in the US health care system in 1991: the cost to the nation, the states and the District of Columbia (unpublished).
4. Wolfe SM, Fugate L, Hulstrand EP, Kamimoto LE, Public Citizen Health Research Group. Worst pills best pills: the older adult's guide to avoiding drug-induced death or illness. Washington, D.C.: Public Citizen Health Research Group, 1988.
5. Cesarean section. In: Wolfe S, Donkin R, Public Citizen Health Research Group. Women's Health Alert. Reading, MA.: Addison-Wesley Publishing, 1990.
6. Kennell J, Klaus M, McGrath S, Robertson S, Hinkley C. Continuous emotional support during labor in a US hospital: a randomized controlled trial. JAMA 1991; 265:2197–2201.

# THE MOUNT SINAI JOURNAL OF MEDICINE

# Forthcoming:

## Grand Rounds

**Bone Marrow Failure Disorders**
Blanche Alter, MD
Mount Sinai School of Medicine

**Is There A Future for Physicians in America?**
Uwe E. Reinhardt, MD
Princeton University

**Osteoporosis: On the Verge of Rational Effective Therapy?**
John Paul Bilezikian, MD
Columbia University College of Physicians and Surgeons

**Obscure Gastrointestinal Bleeding**
Blair S. Lewis, MD
Mount Sinai School of Medicine

**Adrenergic Receptors in Hypertension and Heart Failure**
Clive Rosendorff, MD
University of the Witwatersrand Medical School

**Why Kidneys Fail**
Lance Dworkin, MD
New York University School of Medicine

## Theme Issues

### NOVEMBER 1991

**Asthma: New Concepts**
*Editor:* Zvi M. Marom
plus general articles and index

### MARCH 1992

**Hormones, Metabolism, and Society: In Honor of Rosalind Yalow**
*Editor:* Eugene Straus

### SEPTEMBER 1992

**Human Cancer: From Precurable to Curable**
*Editors:* Max Sung and Takao Ohnuma

### NOVEMBER 1992

**Community Medicine: In Honor of Kurt Deuschle**
*Editor:* Alan Silver

## Available now:

**Toward the Conquest of Pain**
Allan P. Reed, editor
volume 58, no. 3, May 1991
84 pages $15

**Lumbar Disk Disease: Controversies in Neurosurgery**
Leonard I. Malis and Martin H. Savitz, editors
volume 58, no. 2, March 1991
96 pages $15

**The Art and Science of Consultation in Medicine**
Paul L. Gilbert, editor
volume 58, no. 1, January 1991
96 pages $15

*Subscriptions for 1991: $60 individuals, U.S. and Canada, and $65 all libraries; $65 individuals outside U.S. and Canada.*

*To order subscriptions or copies of back issues ($15 each), please send check, payable to The Mount Sinai Journal of Medicine, to the Journal at Box 1094, One Gustave L. Levy Place, New York, NY 10029-6574; counter sales at Suite 4D, 19 E. 98th Street, New York, NY (phone: 212 241-6108; FAX: 212 722-6386).*

LEVY LIBRARY, MT. SINAI MEDICAL CENTER

# THE
# MOUNT SINAI JOURNAL
# OF MEDICINE

OLUME 58 NUMBER 6 NOVEMBER 1991

NOV 29 1991

## CONTENTS

**NEW CONCEPTS IN ASTHMA**
*Zvi M. Marom, editor*

**Introduction: The Asthma Room** *Alvin S. Teirstein* .................... 463

**The Differential Diagnosis of Asthma** *Alvin S. Teirstein* .............. 466

**The Role of Mast Cells in Bronchial Asthma: Mechanisms and Possible Therapeutic Indications** *Zvi M. Marom* .................. 472

**Occupational Airway Diseases** *E. Neil Schachter* ...................... 483

**The Role of Atopy in Asthma** *Kirk Sperber* ............................ 494

**Asthma in Children and Its Management: An Update** *Diana Lowenthal and Meyer Kattan* ........................................ 500

**Pulmonary Function Testing in Bronchial Asthma: Standard and Emerging Techniques** *Lee K. Brown* ............................... 507

**GRAND ROUNDS**

**Bone Marrow Failure Disorders** *Blanche P. Alter* ...................... 521

**THE FUTURE OF MEDICINE**

**Editorial** *Sherman Kupfer* ................................................... 535

**An Adolescent Aspiring Surgeon's Reflections on *Mortal Lessons: Notes on the Art of Surgery* by Richard Selzer** *Andres Schanzer* ................................................................ 536

**Comment** *Michael W. DiGennaro* ........................................ 540

**List of Journal Reviewers in 1991** ........................................ 541

*continued inside*

Veterans Affairs

BETH
ISRAEL
MEDICAL
CENTER

# THE MOUNT SINAI JOURNAL OF MEDICINE

The Mount Sinai Journal of Medicine is published by The Mount Sinai Medical Center of New York and has the following affiliates: Beth Israel Medical Center, New York; Bronx Veterans Affairs Medical Center, New York; and Elmhurst Hospital Center, New York.

**Editor-in-Chief**

Sherman Kupfer, M.D.

**Editor Emeritus**

Lester R. Tuchman, M.D.

**Associate Editors**

Harriet S. Gilbert, M.D. Julius Wolf, M.D.

**Managing Editor**

Claire Sotnick

**Assistant Editors**

David H. Bechhofer, Ph.D.
Constanin A. Bona, M.D., Ph.D.
Edward J. Bottone, Ph.D.
Jurgen Brosius, Ph.D.
Steven T. Brower, M.D.
Lewis Burrows, M.D.
Joseph S. Eisenman, Ph.D.
Richard A. Frieden, M.D.
Steven Fruchtman, M.D.
Paul L. Gilbert, M.D.
James H. Godbold, Ph.D.
Richard S. Haber, M.D.
Noam Harpaz, M.D.
Dennis P. Healy, Ph.D.
Barry W. Jaffin, M.D.
Andrew S. Kaplan, D.D.S.
Samuel Kenan, M.D.
Suzanne Carter Kramer, M.Sc.
Mark G. Lebwohl, M.D.
Kenneth Lieberman, M.D.
Charles Lockwood, M.D.
Steven Markowitz, M.D.
Bernard Mehl, D.P.S.
Myron Miller, M.D.
Edward Raab, M.D.
Allan Reed, M.D.
Allan E. Rubenstein, M.D.
David B. Sachar, M.D.
Henry Sacks, M.D.
Robert Safirstein, M.D.
Ira Sanders, M.D.
Martin H. Savitz, M.D.
Clyde B. Schechter, M.D.
Michael Serby, M.D.
Phyllis Shaw, Ph.D.
George Silvay, M.D.
Barry D. Stimmel, M.D.
Nelson Stone, M.D.
Max Sung, M.D.
Rein Tideiksaar, Ph.D.
Richard P. Wedeen, M.D.
Michael F. Wesson, M.D.

**Editorial Board**

Barry Freedman, M.B.A.
Richard Gorlin, M.D.
Nathan Kase, M.D.
Panayotis G. Katsoyannis, Ph.D.
Charles K. McSherry, M.D.
Jack G. Rabinowitz, M.D.
John W. Rowe, M.D.
Alan L. Schiller, M.D.
Alan L. Silver, M.D.
Alvin S. Teirstein, M.D.

*The Mount Sinai Journal of Medicine* (ISSN No. 0027-2507; USPS 284-860) is published 6 times a year in January, March, May, September, October, and November in one indexed volume by the Committee on Medical Education and Publications (Owner), The Mount Sinai Hospital, Box 1094, One Gustave L. Levy Place, New York, NY 10029-6574. Subscription price: individuals, U.S. and Canada, $60 per year; all libraries, $65; individuals outside the U.S. and Canada, $65. Single copies, $15. New subscriptions begin with the first issue of the calendar year. Send notice of change of address 60 days before the change is effective. Second-class postage paid at New York, NY and at additional mailing offices. POSTMASTER: Send address changes to The Mount Sinai Journal of Medicine, Box 1094, One Gustave L. Levy Place, New York, NY 10029-6574. Copyright 1991 The Mount Sinai Hospital.

HE
OUNT SINAI JOURNAL
F MEDICINE

Volume 58
Number 6
November 1991

## CONTENTS *continued*

**Mount Sinai School of Medicine Faculty Awards and Prizes** ........... 542
**Subject, Author, and Title Index to Volume 58** ............................ 543
**Editorial Note: Death of David A. Dreiling** ............................ cover 3

Index to Advertisers

Eli Lilly and Company 548

# New Concepts in Asthma

## *Introduction: The Asthma Room*

Alvin S. Teirstein, MD

The asthma room, one of the last vestiges of the "snake-pit," is a busy integral part of most hospital emergency suites. "Give epi to number 3." "Hang another aminophylline drip on number 7." "Keep that mask on your face, Mr. Smith!" A room full of gasping, wheezing patients who are subjected to a therapeutic ritual which has not changed significantly in more than a generation. A routine, all too familiar to most of the struggling patients, nurses, and physicians, is repeated nightly. Patients and staff know that they will be reliving this experience again and again in this or some other asthma room. It is no wonder that despite the Madison Avenue depictions in advertisements of pulmonary agents that behave like angels waving their great wings and wafting delicious air in and out of the lungs, and despite an explosion of research in asthma, one of the most treatable diseases in medicine causes more and more deaths each year.

During the past decade several epidemiologic, biologic, and sociologic factors which have enormous impact on the care of patients with asthma have emerged:

1. Death due to asthma is increasing.
2. Asthma is an inflammatory lung disease, not merely bronchospasm.
3. There is no precise definition of asthma.
4. Most patients with asthma who have access to continuous cogent medical care rarely require emergency room treatment or hospitalization.

**Increasing Unexplained Death Rate.** A mounting literature has documented the rise in asthma mortality through the 1980s. Changes in coding, increased hospital-based treatment, increased physician awareness, and abuse of beta-agonist therapy have all been touted as reasons for the reported increase in asthma deaths. In England, Australia, Hong Kong, and the United States we continue to debate the causes of this mounting death rate due to asthma. There is almost no reference to physicians' failure to incorporate the growing new knowledge of the mechanics of the disease or the frequently medieval mode of application of medical care.

**Asthma as an Inflammatory Lung Disease.** Paradoxically, important new advances in understanding the pathogenesis of asthma have been growing in tandem with the unexplained increased death rate. Through the utilization of the flexible fiberoptic bronchoscope to obtain airway constituents, asthma has been firmly established as an inflammatory airways disease. A complicated sequence of events, including receptor response to allergens and other irritants, recruitment of effector cells, reaction to a variety of neuropeptides, cytokines, lipoxygenase products, and leukotrienes, secretion of mucus and impaired mucociliary clearance, microvascular leakage, response of the bronchial epithelial and goblet cells, edema formation, and smooth-muscle reactivity all have been demonstrated in asthma. Clearly, asthma is not merely a bronchospastic disease. Although bronchodilator therapy is an important component of therapy, such treatment alone is as incomplete as treating a fracture only with analgesics or congestive heart failure only with oxygen.

**Imprecision in Definition.** One of the major problems with the study and care of patients with asthma is the absence of a precise definition of the disease. Asthma is defined by the American Thoracic Society as a disease characterized by an increased responsiveness of the trachea and bronchi to various stimuli and manifested by widespread narrowing of the airways that changes in severity either spontaneously or as a result of therapy. Here, again, the emphasis is on bronchoconstriction. This definition would apply to a variety of other entities, including inhalation of toxic gases

---

Alvin S. Teirstein is director of the Division of Pulmonary and Critical Care Medicine, Department of Medicine, and Florette and Ernst Rosenfeld and Joseph Solomon Professor of Medicine, Mount Sinai School of Medicine. Address reprint requests to him at Box 1232, One Gustave L. Levy Place, New York, NY 10029.

and foreign bodies, left ventricular heart failure, pulmonary emboli, and chronic bronchitis.

In its simplest forms, asthma is easily defined as an acute episode or exacerbation of dyspnea accompanied by wheezing, incited by a specific agent or set of circumstances. The patient who enters a home which houses a cat and begins to wheeze, the patient who is exposed to isocyanates in the plastics factory where he is working and wheezes during the work week but recovers at home, the patient who wheezes when jogging or climbing a hill on a windy January day, and the patient who wheezes when the in-laws come for a long visit (one of Freud's first patients was an asthmatic) all have asthma. Unfortunately, most asthmatic patients do not so obviously disclose themselves, or when they do, have in addition episodes of dyspnea that occur with inclement weather or upper respiratory infections, and also at times occur when there is no readily identifiable inciting cause.

The disease most often confused with asthma is chronic bronchitis. To confound the diagnostic problem, the American Thoracic Society has defined chronic bronchitis as cough productive of sputum for at least three months of the year for at least two years. According to that definition, most adult asthmatics have chronic bronchitis. Indeed, in light of the emerging knowledge that asthma is an inflammatory airways disease, asthma should be considered a form of chronic bronchitis.

The ultimate offense perpetrated by the sages of pulmonary medicine has been the introduction of the term "chronic obstructive pulmonary disease" (COPD). In tacit recognition of the difficulty of precisely defining the three major diseases causing physiologically demonstrable obstruction to airflow, asthma, chronic bronchitis, and pulmonary emphysema are now routinely grouped together under that rubric. Hundreds of published articles incorporate that term in their titles, and national and international meetings are held to discuss it.

Unfortunately, this widespread blurring of distinctions has dire therapeutic implications. When the three disorders are lumped together as COPD, the physician is relieved of the burden of establishing a precise diagnosis and all patients receive the same treatment—bronchodilators, oxygen, adrenocorticosteroids, and antibiotics—and the same follow-up. However, asthma, bronchitis, and emphysema are not the same disease:

- The smoking, coughing, and wheezing patient producing green and yellow sputum is easily recognized as having chronic bronchitis. This patient has frequent bacterial infections, and interdiction of smoking and use of antibiotics is effective therapy.
- The nonsmoking asthmatic often has attacks precipitated by viral infections. Although such a patient may produce "purulent" sputum, excellent studies have shown that respiratory viruses are the inciting infectious agents in asthma, whereas bacterial infection is unusual. Antibiotics are rarely indicated. The asthmatic patient requires bronchodilators but must also receive therapy directed at the inflammatory bronchial disease and therapy directed at prevention of future attacks.
- Unfortunately, patients with pulmonary emphysema suffer from a destructive lung disease with airflow obstruction and little, or no, inflammation. Despite numerous conflicting studies of the efficacy of antibiotics, bronchodilators, and adrenocorticosteroid medications, the mainstays of treatment for pure pulmonary emphysema are bronchodilators, low-flow oxygen, treatment of cor pulmonale, and rehabilitation.

Thus, the widespread practice of grouping patients with asthma, chronic bronchitis, and emphysema together under the umbrella of COPD is sloppy thinking that leads to thoughtless, often harmful therapy.

**Importance of Continuous Medical Care.** Asthma is one of the most treatable medical diseases. A superb array of relatively risk-free medications is now available. Almost all asthmatics can function well with a combination of hand-held meter dose inhalers tailored to the patient's specific needs. Most important is a close relationship between doctor and patient. The majority of patients with uncontrolled asthma are either not under the care of a knowledgeable physician who knows them well, or are self-medicating. For example, prescribing a packet of adrenocorticosteroid tablets with a preset diminishing dosage schedule frequently assures another attack of asthma a few days after the medication has been discontinued. The value of a close doctor-patient relationship in the treatment of asthma is evidenced by the patients populating emergency rooms and their inevitable extension, hospital beds. Very few patients with ongoing treating physicians must resort to hospital emergency rooms for treatment, and even fewer require hospitalization. Pilot studies have clearly demonstrated that millions of healthcare dollars could be saved if all patients with asthma had 24-hour access to competent medical care.

Currently, asthma, an old disease, is being viewed in a new way. This new look has arrived coincident with increasing asthma mortality. Effective treatment is available and requires an understanding that asthma is an inflammatory disease. While bronchodilator therapy is effective for relief of symptoms, environmental manipulation and the utilization of available measures to prevent and relieve airways inflammation are essential for optimal therapy. Many means are currently available to achieve this goal; with the rapidly growing knowledge of the pathogenesis of airways inflammation, new and precise therapies are in the offing. A measure of their success will be the abolition of the hospital asthma room.

# The Differential Diagnosis of Asthma

ALVIN S. TEIRSTEIN, M.D.

IT IS most difficult to present the differential diagnosis of a disease that has no universally accepted definition. Even a much-heralded CIBA symposium failed to produce a precise definition (1). Vaughan et al. (2) have proposed this definition: "a disorder characterized by reversible bronchospasm with wheezing, the symptoms occurring in short-lived paroxysms." Reed (3) has asked, "Is asthma a single disease with a variety of provocative factors and variable course, or is asthma a syndrome with a variety of differing etiologies that have different courses?" The introduction of the term "chronic obstructive pulmonary disease" (COPD) has impeded further attempts at precise definition.

It is not surprising, therefore, that many physicians and young housestaff officers no longer precisely use the term asthma, opting instead for COPD. Hence, almost every patient who wheezes is treated with antibiotics, bronchodilators, and adrenocorticosteroids, regardless of their disease. Consider what happened to these patients:

A 60-year-old male cigarette smoker was admitted to the emergency room with sudden onset of acute respiratory distress, wheezing, hypoxemia, and moderate hypercapnia. He was sent to the "asthma room," treated with intravenous theophylline, hydration, and inhalation of beta adrenergic medications. He failed to improve, and subcutaneous epinephrine was administered. Subsequently, a chest x-ray revealed pulmonary edema, and his electrocardiogram showed an acute myocardial infarction. The incorrect diagnosis of asthma or COPD could have been fatal.

A 58-year-old woman consulted her family physician because of gradually increasing dyspnea and wheezing. She was treated with bronchodilators. Her symptoms worsened. A second physician observed that her noisy breathing was localized to the right upper chest. The remainder of her lung fields were clear. Chest radiographs demonstrated obstructive emphysema of the right upper lobe, and bronchoscopy showed a fungating mass almost completely obstructing the right upper lobe bronchus.

These two patients represent common, everyday presentations which demand a precise differential diagnosis of asthma.

Asthma can best be defined as attacks of dyspnea and wheezing, occurring under predictable circumstances, which improve spontaneously or with therapy. The key terms are *wheezing* and *reversibility*. Most adult asthmatics have a history of previous similar bouts. Although atopy is frequent, classical IgE-mediated attacks are less common in adults than children. Most often the adult patient reports that wheezing occurs with exercise, upper respiratory infections, cold or humid weather, exposure to fumes and dust, and emotional stress.

In the absence of overt wheezing, the diagnosis of asthma in a patient with a history of dyspnea and cough is difficult (4). Suspicion is aroused because the patient's symptoms occur under circumstances often associated with asthma—for example, exercise, nocturnal cough, dyspnea. When the patient is examined, the clinician hears no wheezes. Even spirometry and flow-volume loops may be normal. In such a patient, bronchial provocation testing with methacholine, cold air, or histamine may be diagnostic of bronchial hyperactivity. Indeed, in the absence of a specific definition for asthma, some authorities have proposed that a positive bronchial challenge test is diagnostic of asthma (5, 6).

Classically, asthma has been divided into extrinsic and intrinsic groups. Extrinsic asthma occurs in atopic individuals with a strong family

---

From the Div. of Pulmonary and Critical Care Medicine, Dept. of Medicine, Mount Sinai School of Medicine. Address reprint requests to the author, director of the Div. of Pulmonary and Critical Care Medicine and Prof. of Medicine, Box 1232, One Gustave L. Levy Place, New York, NY 10029.

history of asthma. Attacks are often seasonal or invoked by exposure to known allergens. Peripheral blood eosinophilia and elevated serum IgE levels are common. Sputum smears reveal many eosinophils. Patients with intrinsic asthma tend to be older, may or may not have a history of atopy, and are prone to attacks throughout the year. There is a large overlap of these two groups. Adults frequently give a history of seasonal extrinsic asthma in childhood which abated in their teens, only to reoccur later in life with a pattern of attacks suggesting intrinsic asthma. It is now generally agreed that this artificial separation into two groups has little or no clinical relevance and the distinction has been discarded by most modern authorities.

Wheezes and rhonchi are the clinical hallmarks of asthma. Wheezing is absent in the asthmatic only at the two clinical extremes of the disease, when the patient is well or when the airways are so obstructed that almost no air can pass and the velocity of air flow is minimal. Wheezes have been subjected to acoustic analysis. True wheezes are continuous, high pitched, and heterogeneous and have a frequency of approximately 400 Hz. They are velocity dependent and occur when the air space walls almost touch. Rhonchi, which also occur in asthma, are low pitched, often continuous, may clear with coughing, and have a frequency of about 200 Hz (7, 8). Both airway and extra-airway forces produce wheezing (Fig.).

In asthma, airway narrowing occurs as a complicated sequence of events including release of mediators from effector cells, cellular inflammation-infiltration, bronchial wall edema, hypersecretion of mucus, and smooth muscle contraction. In the increasing number of patients who die during an asthmatic attack, the airways are edematous, inflamed, and plugged with mucus casts.

Standard pulmonary function tests can be helpful in the differential diagnosis of asthma. Almost all patients who wheeze will exhibit classic obstructive spirometric dysfunction, that is, slight reduction in forced vital capacity (FVC), greater reduction in the forced expiratory volume in one second ($FEV_1$), decreased $FEV_1/FVC$ ratio, and slow flow-rates. Since reversibility is typical of asthma, a prompt improvement in spirometric parameters after administration of inhaled beta agonists is expected. When in doubt, a bronchial challenge test is an excellent diagnostic aid. Chest radiographs are usually normal in patients with asthma, but may show hyperlucent lung fields with flattened diaphragm or atelectasis of a segment, a lobe, or even an entire lung due to mucus (impaction), particularly in patients with asthma and bronchopulmonary aspergillosis.

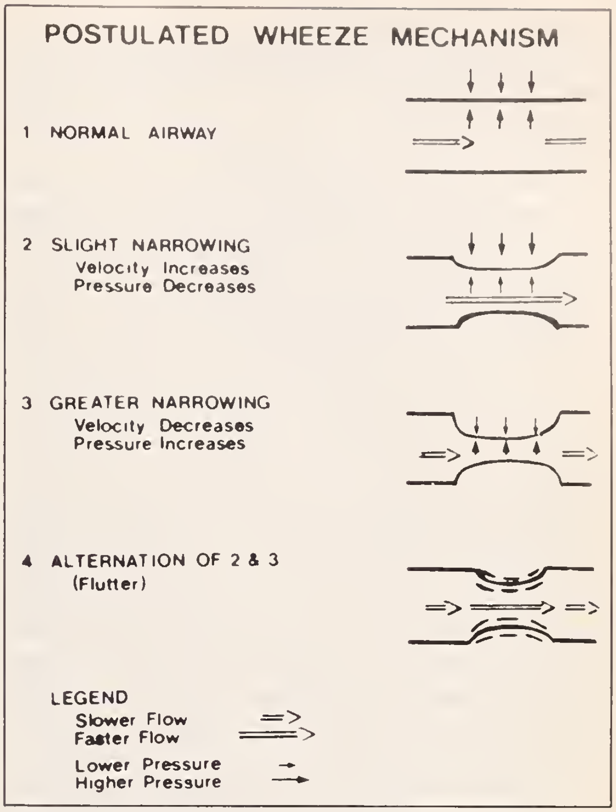


Fig. Stability of airway wall depends on balance between internal air pressure and external forces and on mechanical characteristics of airway itself (1). When lumen narrows, air velocity must increase through constricted region to maintain constant mass flow rate. According to Bernouilli's principle, this leads to decrease in air pressure in constricted region, thus allowing external compressive forces to further collapse airway (2). When lumen has been reduced so much that flow rate decreases, process begins to reverse itself as pressure inside airway begins to increase and reopen lumen (3). When conditions are right, the airway wall "flutters" between open and nearly closed states and produces a continuous sound whose amplitude, pitch, and duration depend on airflow and mechanical parameters involved (4). *Reprinted* with permission from Murphy RLH, Holford SK. Lung sounds. ATS News 1980; 8(4):24–29.

Although the onset of asthma in adults is common, particularly after severe viral infections such as influenza, the onset of wheezing in an adult mandates that the physician consider causes of wheezing other than asthma. The diseases most often confused with asthma are:

- cardiogenic pulmonary edema
- chronic bronchitis
- pulmonary emphysema
- upper airway obstruction
- pulmonary emboli
- aspiration of gastric contents
- miscellaneous nonasthmatic causes of wheezing (carcinoid tumor, eosinophilic pulmonary infiltrates)
- factitious asthma.

One of the oldest and most ignored aphorisms in medicine is "All that wheezes is not asthma." Despite this truism, every one of these diseases has been, and continues to be, misdiagnosed as asthma.

**Cardiogenic Pulmonary Edema.** Although patients with noncardiogenic pulmonary edema may wheeze, wheezes are much more common in patients with left ventricular cardiac failure. The airways are narrowed due to interstitial edema and transudation of fluid into the airways. The patients usually have a history of heart disease or an acute episode indicating myocardial damage or arrhythmia. Many are cigarette smokers with chronic cough and sputum production, which may compound the diagnostic problem. Acute therapeutic trials are often valuable in differentiating cardiac from true asthma. The patient with pulmonary edema will not improve with bronchodilators but should improve with a good cardiotonic regimen. Chest radiographs, electrocardiograms, and in some patients, right side of the heart catheterization will establish the cardiac etiology of wheezing in left ventricular failure.

**Chronic Bronchitis.** The most difficult disease to differentiate from asthma is chronic bronchitis. This diagnostic dilemma has given rise to such terms as "asthmatic bronchitis" and the ultimate in diagnostic obfuscation, "chronic obstructive pulmonary disease" (COPD). Chronic bronchitis is defined clinically. According to the American Thoracic Society diagnostic standards, chronic bronchitis is a disease characterized by cough and sputum production for three months of the year, for at least two years (9, 10). This imprecise definition obviously would include other diseases such as bronchiectasis, cystic fibrosis and, of course, asthma. This is not surprising since asthma is a state of chronic bronchial inflammation with acute exacerbations, a special form of chronic bronchitis. Patients with classical chronic bronchitis tend to be older cigarette smokers with chronic cough productive of mucoid and mucopurulent sputum. They usually do not have a history of atopy, eosinophilia, or elevated serum IgE levels. Sputum examination reveals many polymorphonuclear leukocytes, no increase in eosinophils, and bacteria.

Patients with chronic bronchitis have persistent symptoms exacerbated by recurring respiratory infections. With progression of the disease, ventilation-perfusion mismatches lead to chronic hypoxemia, secondary polycythemia, and eventually pulmonary hypertension, cor pulmonale, and right heart failure. The chronic bronchial inflammation causes edema, goblet-cell hyperplasia, and the presence of thick mucus which obstruct the airways. Therefore, wheezes and rhonchi are common. Bronchospasm does occur and can respond to the usual bronchodilators, causing confusion with asthma. In some patients it is impossible to definitively differentiate chronic bronchitis from asthma. In the older cigarette smoker with chronic cough productive of thick sputum, the diagnosis is usually chronic bronchitis, not asthma.

**Pulmonary Emphysema.** Pulmonary emphysema is a pathologic diagnosis characterized by enlargement of the distal airways and acini with interruption of the alveolar walls and enlargement of the alveoli. The lungs are usually large with an increased total lung capacity. In common with all diseases of the airways, pulmonary function studies demonstrate obstructive dysfunction. In contrast to asthma and bronchitis, the single-breath carbon monoxide diffusing capacity is usually decreased in emphysema. The disease occurs almost exclusively in cigarette smokers and progresses slowly but relentlessly. Physical examination reveals an elderly patient with a large thorax on an otherwise wasted frame. Chest wall and diaphragmatic motion is decreased. The clinical sine qua non is markedly decreased breath sounds. In uncomplicated pulmonary emphysema wheezes are audible only at the end of forced expiration, probably as a result of collapse of the airways by the increased intrathoracic pressure acting on obstructed dilated distal air spaces. Standard chest radiographs are a poor technique for the diagnosis of pulmonary emphysema except when bullae are present. The radiologist is on firm ground only when the radiograph reveals large hyperlucent lungs, flattened diaphragm, small heart, and large pulmonary arteries. Although a rare patient with asthma may develop pulmonary emphysema, there is little excuse for confusing these two diseases.

**Upper Airway Obstruction.** Wheezing patients with disease involving the larynx, trachea, and main bronchi invariably are misdiagnosed as having asthma by the first treating physician. The onset is usually insidious with a prolonged period of dyspnea and wheezing with exertion. The clinician hears wheezes and prescribes bronchodilators. It is many months, sometimes years, until the telltale audible stridor and palpable rhonchus is detected by the last physician, usually a consultant or hospital admitting doctor. Little do these latter-day geniuses appreciate the subtlety of the early signs and symptoms of upper airway obstruction.

Common causes of obstruction to the major airways range from bilateral paralysis of the vocal cords to an aspirated foreign body lodged in a major bronchus (Table). Even when the obstructing lesion compromises air flow in a lobar bronchus, the presence of a wheeze or rhonchus may provoke an erroneous diagnosis of asthma (11–13).

The clinical hallmarks of upper airway obstruction are stridor or a localized wheeze or rhonchus. Obstruction of the larynx or trachea produces a continuous, sonorous, low-pitched (less than 200 Hz) sound, often loudest on inspiration. Palpation reveals course vibrations which often are of great aid in localizing the site of obstruction. Occasionally, narrowing of a main bronchus or lobar bronchus may emit rhonchi which are transmitted to the trachea. Careful auscultation of the lung fields reveals normal breath sounds with no wheezes in the more peripheral airways. Standard chest radiographs may demonstrate partial atelectasis or obstructive emphysema of a single lobe or entire lung and/or mediastinal or thyroid masses. On occasion the standard inspiratory posterior-anterior radiographs are normal, particularly in patients with tracheal or laryngeal lesions. When the obstruction is distal to the carina, *expiratory* posterior-anterior radiographs may demonstrate obstructive emphysema of the lung or lobe downstream from the compromising lesion.

Upper airway obstruction can be elegantly demonstrated by performance of pulmonary function tests. The maximum expiratory flow rate is severely limited and the flow-volume loop gives a typical squared-off pattern resembling a rectangle. The shape of the curve may indicate whether the obstruction is extrathoracic or intrathoracic (13).

**Pulmonary Emboli.** Patients with acute pulmonary emboli may present with wheezing (14). Since pulmonary emboli are common in patients with heart disease, the adventitious sounds may be secondary to left ventricular failure and cardiogenic pulmonary edema precipitated by the pulmonary emboli. Wheezing also occurs in patients with pulmonary emboli and no primary cardiac disease. The wheezes may be heard diffusely, not necessarily localized to the area of pulmonary infarction. Indeed, there may be no clinically detectable infarcted lung. It has been proposed that bronchospasm is induced by bronchoactive substances liberated by the blood clots lodged in the pulmonary vasculature. Release of bronchoconstricting agents such as serotonin, bradykinin, platelet-activating factor and histamine have all been implicated as causes of wheezing after pulmonary embolization.

The clinical setting of a patient at risk for pulmonary emboli—cardiac disease, pregnancy, postoperative state, or any debilitating disease—and the absence of a prior history of asthma should alert the physician that the patient's dyspnea and wheezing is not due to asthma. Prompt employment of a variety of studies, including chest radiographs, radionuclide ventilation and perfusion lung scans, tests of venous flow in the lower extremities, and, when necessary, pulmonary artery angiography, should establish the correct diagnosis of pulmonary emboli as the cause of the patient's wheezing.

TABLE
*Causes of Upper Airway Obstruction*

| Site | Cause |
|---|---|
| Larynx | Edema, epiglottitis, bilateral vocal cord paralysis, tumors (benign and malignant) |
| Trachea | Stenosis (common postendotracheal intubation and tracheostomy), primary tracheal tumors (benign and malignant), thyroid tumors, metastatic tumors, tracheomalacia, granulomas, papillomas, amyloid |
| Main bronchi | Primary lung cancers, metastatic cancers, foreign bodies, granulomas, amyloid, lymphomas |

**Aspiration of Gastric Contents.** Regurgitation with aspiration of partially digested food and hydrochloric acid from the stomach or oral bacterial flora can result in a variety of pulmonary syndromes (15–17). Lung abscess and bronchopneumonia are usually not clinically characterized by wheezing. Aspiration of foreign bodies (partially digested food) may lead to localized wheezes at the sites of impaction of the aspirated particles. In younger patients with significant amounts of low pH gastric juice, an acute diffuse bronchial inflammation may rapidly occur, followed in a few hours by pulmonary edema (Mendelson's syndrome) (18). During the initial stages of bronchial edema, wheezing, cough, and dyspnea are common and may be misdiagnosed as asthma. The history of aspiration and the absence of prior asthma should alert the physician to the correct diagnosis of aspiration pneumonitis.

True asthma may be common in patients with esophagogastric reflux. Several authors have proposed that upper gastrointestinal acid reflux commonly causes asthma, either due to direct acid lung injury or by inducing reflux bronchospasm. With treatment of the acid reflux syndrome, the asthma may clear (19, 20).

**Miscellaneous Nonasthmatic Wheezing.** Almost any pulmonary disease may cause wheezing. Wheezes and rhonchi are common in bacterial and viral pneumonias. The clinical presentation and radiographic abnormalities should preclude confusion with asthma. It must be emphasized, however, that pulmonary infection, as well as upper respiratory infections, may precipitate or exacerbate attacks of true asthma. The first episode of asthma in the adult may occur during or as a sequela of pneumonia. Inhalation of toxic fumes such as nitrous oxide, sulfur dioxide, chlorine, formaldehyde, and especially smoke inhalation can result in severe asthmalike attacks of wheezing. Bronchial edema and detritus with wheezing and dyspnea often herald the first clinical stage of smoke inhalation. This may be followed in a few hours by capillary leakage and noncardiogenic pulmonary edema (21). The history of toxic inhalation precludes confusion with asthma. Rarely, the initial symptoms of a patient with carcinoid syndrome may be attacks of asthma (22). Interestingly, most of these patients have their primary carcinoid tumor below the diaphragm, and not in the thorax. It is assumed that serotonin secreted by the carcinoid tumor triggers bronchospasm. Hemoptysis, atelectasis, gastrointestinal complaints, flushing and tachycardia, evidence of an endobronchial tumor, or the vasoactive effects of serotonin accompanying wheezing may alert the physician to the presence of a carcinoid tumor. Appropriate biochemical and radiographic studies should yield the correct diagnosis.

Wheezing which occurs in the so-called pulmonary vasculitides (23), which usually include periarteritis nodosum (PAN), Wegener's granulomatosis, and Churg-Strauss disease, need special mention. An ancient clinical axiom states that the onset of asthma in an adult patient should provoke a search for PAN. Nonsense! Although the lungs may participate in the systemic vasculitis, pulmonary symptoms occur in the minority of patients and wheezing is rare in PAN. Pulmonary involvement is common in Wegener's granulomatosis. Pulmonary masses which may cavitate, or bronchial obstruction by necrotic granulomas with pathologic evidence of vasculitis, are common in Wegener's granulomatosis. Wheezing is rare. Asthma is almost universal in patients with Churg-Strauss syndrome (24). This disease is an allergic granulomatosis marked by the presence of peripheral eosinophilia and true asthma. Similarly, the hypereosinophilic syndrome and any "allergic" disease in a patient with markedly elevated blood levels of eosinophils, such as tropical eosinophilia and Loeffler's pneumonia, may initially appear as asthma. The multisystem involvement and radiographic abnormalities should lead the physician to the correct diagnosis in patients with vasculitis and eosinophilic syndromes.

**Factitious Asthma.** It is not uncommon to encounter a patient complaining of dyspnea and wheezing who has been treated with the entire gamut of therapy available for asthma but has had no salutary response. In most of these patients the history fails to reveal atopy or worsening of symptoms due to all but one of the usual provocations—infections, inhalations of dust and fumes, cold or humid weather, exercise; the exception is psychological stimuli. A careful physical examination reveals that the audible wheezes, often heard across the room, originate in the throat and are produced by exhaling against constricted laryngeal muscles (25). Except when a pathologic lesion exists in the upper airway, the correct diagnosis is factitious asthma and is supported by a negative methacholine challenge test.

**Summary.** The clinical hallmarks of asthma are wheezing and reversibility. Any disease that impairs air flow through obstructed airways may cause wheezing. Patients with true asthma may give a history of allergy and past attacks of dyspnea and wheezing occurring when exposed to allergens, inhaled irritants, upper respiratory infection, cold and humid air, exercise, and emotional stress. When encountering a wheezing dyspneic patient who does not report such a history, it behooves the physician to entertain the possibility that the patient may have a disease other than asthma. Chronic bronchitis, pulmonary emphysema, cardiogenic pulmonary edema, pulmonary emboli, aspiration of gastric contents, and upper airway obstruction are the common causes of nonasthmatic wheezing. In almost every instance a wide spectrum of easily obtainable data, particularly historical, are available to alert the physician that the patient's dyspnea and wheezing are not due to asthma. Laboratory data are also readily available to buttress the correct diagnosis.

### References

1. Fletcher CM, Gilson JG, Hugh-Jones P, et al. Terminology, definitions and classification of chronic pulmonary emphysema and related conditions: a report of the conclusions of a CIBA guest symposium. Thorax 1959; 14:286–299.
2. Vaughan JH, Eng MT, Mathison DA, Stevenson DD, Berman SZ, Braude AI. Bronchial asthma: pathophysiology. In: Austen KF, Lichenstein LM, eds. Asthma: physiology, immunopharmacology and treatment. New York: Academic Press, 1973:1–13.
3. Reed CE. Epidemiology and natural history. In: Stein M, ed. New directions in asthma. Chicago, Ill.: American College of Chest Physicians, 1975:291–300.

4. Corrao W, Braman SS, Irwin RS. Chronic cough as the sole presenting manifestation of bronchial asthma. N Engl J Med 1979; 300:633–637.
5. Myers JR, Corrao WM, Braman SS. Clinical applicability of a methacholine inhalation challenge. JAMA 1981; 246:225–229.
6. Pratten MR, Irwin PS. The clinical value of pharmacologic bronchoprovocation challenge. Chest 1984; 85:260–265.
7. Hollingsworth HM. Wheezing and stridor. In: Braman SS, ed. Clinical chest medicine: pulmonary signs and symptoms, vol. 8, no. 2. Philadelphia: WB Saunders–Harcourt Brace Jovanovich, 1989:231–240.
8. Loudon RG, Murphy RLH, Jr. Lung sounds. Am Rev Respir Dis 1984; 130:663–673.
9. ACCP-ATS Joint Committee on Pulmonary Nomenclature. Pulmonary terms and symbols. Chest 1975; 67:583–593.
10. Meneely GR, et al. Definition and classification of chronic bronchitis, asthma and pulmonary emphysema. A statement by the American Thoracic Society, Committee on Diagnostic Standards for Non-tuberculous Respiratory Disease. Am Rev Respir Dis 1962; 85:762–769.
11. Braman SS, Whitcomb ME. Endobronchial metastasis. Arch Int Med 1975; 135:543–545.
12. Acres JC, Kryger MH. Upper airway obstruction. Chest 1981; 80:207–211.
13. Kryger M, Bode F, Antic R, et al. Diagnosis of obstruction of the upper and central airways. Am J Med 1976; 61:85–93.
14. Windebank WJ, Boyd G, Moran F. Pulmonary thromboembolism presenting as asthma. Br Med J 1973; 1:90–94.
15. Bartlett JG, Gorbach SL. The triple threat of aspiration pneumonia. Chest 1975; 68:560–566.
16. Bynum LJ, Pierce AK. Pulmonary aspiration of gastric contents. Am Rev Respir Dis 1976; 114:1129–1136.
17. Schwartz DJ, Wynne JW, Gibbs CP, et al. The pulmonary consequences of aspiration of gastric contents at pH values greater than 2.5. Am Rev Respir Dis 1980; 121:119–126.
18. Mendelson CL. The aspiration of stomach contents into the lungs during obstetric anesthesia. Am J Obstet Gynecol 1946; 52:191–205.
19. Harper PC, Bergner A, Kaye MD. Antireflux treatment for asthma: improvement in patients with associated gastroesophageal reflux. Arch Intern Med 1987;147:56–60.
20. Sontag SJ, O'Connell S, Khandelwal S, et al. Most asthmatics have gastroesophageal reflux with or without bronchodilator therapy. Gastroenterology 1990;99:613–620.
21. Haponik EF, Summer WR. Respiratory complications in burned patients: pathogenesis and spectrum of inhalation injury. J Crit Care 1987; 2:49–74.
22. Todd TR, Cooper JD, Weissberg D, et al. Bronchial carcinoid tumors: 20 years experience. J Thorac Cardiovasc Surg 1980; 79:532–536.
23. Leavitt RY, Fauci AS. Pulmonary vasculitis. Am Rev Respir Dis 1986; 134:149–166.
24. Churg J, Strauss L. Allergic granulomatosis, allergic angitis and periarteritis nodosa. Am J Pathol 1951; 27:277–294.
25. Downing ET, Braman SS, Fox MJ, et al. Factitious asthma: physiological approach to diagnosis. JAMA 1982; 248:2878–2881.

# The Role of Mast Cells in Bronchial Asthma:

## Mechanisms and Possible Therapeutic Implications

Zvi M. Marom, M.D., M.Sc.

THE CLINICAL manifestation of allergic diseases, which in part are mediated by IgE, depends on the actions of different mediators, many derived from mast cells. This brief overview of mast cells and some of their mediators focusses on current concepts of the response of human lungs to mast-cell mediators and clinical implications.

### Mast Cells

In human tissues the mast cells are relatively abundant in skin, thymus, uterus, lymphoid tissue, urinary bladder, tongue, synovia, mesentery, around large and small blood vessels, in the subserosal and submucosal layers of the digestive tract, and in the respiratory tract (1, 2). Mast cells in the lung are found beneath the basement membranes of airways, near blood vessels in the submucosa, adjacent to submucosal glands, scattered throughout the muscle bundles, in the intra-alveolar septa, and in the bronchial lumen. In the human lung, these cells are present in concentrations of 1 to 7 × $10^6$ cells/g lung tissue (3) and may constitute 2% of alveolar tissue (4). Although the mast cell is ubiquitous and much studied, the precursor from which it is derived has never been clearly ascertained.

The characteristic feature of this cell is the presence of numerous dense, metachromatic-staining cytoplasmic granules that may obscure the ovoid-shaped mononuclear nucleus. Heparin is the constituent that imparts the metachromatic staining property to the mast cell. Chloroacetate ester stains are selective for mast cells and neutrophils in tissue sections.

Mast cells play an important role in allergic reactions. The immunologic trigger responsible for mast-cell degranulation is an immunoglobulin of the $\epsilon$ (epsilon) class. IgE is a typical monomeric immunoglobulin, consisting of two heavy and two light chains, that binds to mast-cell and basophil membrane receptors via its Fc portion. The $F(ab)_2$ fragment functions as the recognition site for specific antigenic binding. The bridging of adjacent IgE molecules by specific antigens (usually allergens) or by anti-IgE antibody initiates membrane events that ultimately culminate in the degranulation process.

The physiochemical characteristics of the IgE receptor on rat mast cells and leukemic basophils have been critically analyzed and reviewed (5). The receptor in situ is unclustered, mobile, and univalent, and its aggregation into dimers and higher oligomers triggers degranulation. The receptor is a glycoprotein (6) composed of two subunits, an alpha chain (molecular weight 50,000 daltons) and beta chain (molecular weight 30,000 daltons). Both these chains are composed of at least two domains, in the alpha chain labeled $\text{alpha}_1$ and $\text{alpha}_2$, in the beta chain $\text{beta}_1$ and $\text{beta}_2$ (5). There are 1 to 5 × $10^5$ receptors/mast cell, of which about 10% are occupied in vivo (2). Nonetheless, the extremely high binding constant of IgE to its receptors, $K_A$ $10^9$ to $10^{12}M^{-1}$, combined with the long residence time of approximately 6 weeks, permits the mast cell to concentrate IgE antibodies on the cell surface.

---

From the Division of Pulmonary and Critical Care Medicine, Mount Sinai Medical Center, New York, NY. Address correspondence and reprint requests to the author, Division of Pulmonary and Critical Care Medicine, Box 1232, Mount Sinai Medical Center, One Gustave L. Levy Place, New York, N.Y. 10029.

The physiologic consequences of exposing IgE-sensitized mast cells to antigens against which the IgE molecule is directed results in the secretion of mast-cell granules from which the mediators of allergy and anaphylaxis are derived. The bridging of two IgE molecules by antigen initiates a biochemical cascade involving the movement of calcium into the cell, activation of a cell membrane-associated serine esterase, glycolysis, calcium liberation, microtubule assembly, and sequential exocytosis of the granule (7). Bridging of IgE receptors on rat mast-cell plasma membranes induces the metabolism of arachidonate, the formation of prostaglandin $D_2$, leukotriene $C_4$, and 5-HETE, and a monophasic increase in cyclic AMP.

MacGlashan et al. have developed a technique to isolate human lung mast cells at more than 98% purity (8). Because only rat mast cells were previously available in pure form, most of the information cited above comes from studies with this cell type. However, now that human mast cells are available in relatively pure form, human mast-cell mediator release mechanisms have been investigated and found to be quite similar to that of the rat's mast cells.

The biochemical events initiated by the aggregation of IgE receptors on mast-cell membranes result in a unique degranulation process termed sequential exocytosis (9), which is completed within 1 to 2 minutes at 37°C (2). Studies by Sullivan suggest that a phospholipid diacylglycerol pathway is involved in degranulation of rat mast cells through the formation of potent biochemical fusogens necessary for the fusion of the granule membranes, creating cytoplasmic channels (10). These channels are geographically within the confines of the cell but freely communicate with the bathing media. Thus, granules move into contact with the external environment but not immediately away from the mast cell. In the rat and possibly in humans the granule itself remains intact and is gradually extruded from the cell. Within 30 to 60 minutes granules are evident at some considerable distance from the parent cell (11).

Mast cells can be of two types: mucosal and connective tissue. They stain differently and differ in size (12). There are also mast-cell subtypes in humans which can be detected by the use of monoclonal antibodies. All human mast cells contain tryptase (a tryptic protease) but only connective tissue mast cells (found in skin and to a lesser degree in the airways) contains chymase, a chymotryptic protease.

TABLE 1
*Mast-Cell-Derived Mediators*

**Preformed rapidly eluted under physiologic conditions**
Histamine
Eosinophil chemotactic factors of anaphylaxis (ECF-A)
Intermediate weight ECF-A*
Neutrophil chemotactic factors (NCF)
Superoxide anions
Exoglycosidases (β-dexosaminidase, β-D-galactosidase*, β-glucuronidase)
Serotonin*
Kininogenase
Arylsulfatase A

**Secondary or newly generated mediators**
Slow-reacting substances (SRS-A): LTC, LTD, LTE
Prostaglandins
Monohydroxyeicosatetraenoic acids (HETEs)
Hydroperoxyeicosatetraenoic acids (HPETEs)
Thromboxanes
Platelet activating factor (PAF)*
Prostaglandin generating factor of anaphylaxis (PGF-A)

**Preformed granule-associated mediators**
Heparin
Chymotrypsin/trypsin
Peroxidase*
Superoxide dismutase*
Arylsulfatase B
Inflammatory factors of anaphylaxis (IFA)*

* Demonstrated in mast cells of nonhuman species.

Mast-cell growth and differentiation depend on several mediators. IL-3, which is produced by T-lymphocytes, is a known stem-cell growth factor that can stimulate mouse bone marrow cells to differentiate into the mucosal mast-cell phenotype. IL-4 potentiates this effect. Both are involved as mast-cell growth-promoting cytokines in humans. Another mast cell regulatory cytokine is transforming growth factor β, which is a potent inhibitor of mucosal mast-cell growth and differentiation (13).

## Mast-Cell-Dependent Mediators

The secretory reaction in the mast cell leads to granule discharge, from which the mediators of allergy are derived. The primary source of information on mediators has been the rat peritoneal mast cell, although considerable information is known about the mediators from human lung, skin, and nasal mast cells, the guinea pig lung mast cell, and the human basophil.

The rat mast-cell granule provides three distinct sources of mediators (Table 1): (a) preformed mediators, contained in the granule matrix and readily displaced by physiologic concentrations of salt such as are found in tissue fluid; (b) secondarily formed mediators generated by the interaction of primary mediators and nearby cells and tissues; and (c) granule matrix-derived mediators, which are preformed but do not readily dissociate from the granule after discharge.

The physiochemical nature of the interaction of mediators with the mast-cell granule matrix provides a temporal sequence of mediator dissociation from the granule. Preformed mediators such as histamine are released into the tissue fluid immediately. Several of the secondary mediators are formed immediately (prostaglandins) or within minutes (slow-reacting substances, SRS). The granule matrix and its associated mediator remain in the tissue for hours until dissolution occurs or they are phagocytosed.

The events initiated by the mediators depend on the tissues into which they are released. Thus histamine in the skin and nose may produce vasopermeability primarily; in the lung histamine may also induce bronchospasm. Mediator discharge clearly initiates the events associated with immediate hypersensitivity, which are apparent within minutes and are explosive. Late-phase allergic reactions occur in response to granule-derived mediators as well. Thus, 2 to 8 hours after degranulation there is an influx of polymorphonuclear leukocytes followed by mononuclear cells over the subsequent 24 hours.

There are some other important immunologic triggers of mast-cell secretion besides the IgE antigen. These include a family of histamine-releasing factors (14), IL-1 and IL-3. Nonimmunologic stimuli can also activate human mast cells to release their mediators. These include substance P; the complement fragments C3a, C4a, and C5a; ADP; and drugs such as opiates. These non-IgE-mediated stimulators are more effective on connective tissue and cutaneous mast cells than on mucosal mast cells.

The growing appreciation of the diverse consequences of mast-cell degranulation has led to an expanding list of mediators produced by this event. It is beyond the scope of this review to extensively discuss each mediator individually. Rather, a brief description of representative human mast-cell mediators is given, with special focus on their physiologically important actions.

**Histamine.** The main depot of tissue histamine is in the mast cell, and under physiologic conditions the tissue content of histamine corresponds closely to the number of mast cells (2). Histamine is a β-imidazolylethylamine (molecular weight 111) formed by decarboxylation of L-histidine.

Two cellular receptors for histamine, termed $H_1$ and $H_2$ receptors, mediate the biologic effects of histamine after mast-cell degranulation. Stimulation of $H_1$ receptors may result in smooth muscle contraction, increased vascular permeability, pruritus, prostaglandin generation, and increases in cyclic GMP (15). $H_2$ receptor stimulation increases gastric acid secretion, cyclic AMP production, and mucus secretion in the lung and inhibits lymphokine release, neutrophil enzyme release, eosinophil migration, and T-lymphocyte-mediated cytotoxicity (15). Vasodilation (hypotension), flush, and headache result from stimulation of both $H_1$ and $H_2$ receptors. Since anaphylaxis is associated with hypotension, the treatment of this condition requires the use of both $H_1$- and $H_2$-receptor antagonism.

Histamine is assayed by either the old mode, fluorometric assay, which is sensitive to 1–5 ng/mL, or the new radioimmunoassay, which is sensitive to 10–50 pg/mL. Histamine can be measured in plasma (normal, <500 pg/mL; average, 270 pg/mL), in urine (normal, <25 ng/24 hr), and in lung secretions obtained by bronchoalveolar lavage, where it is increased in asthmatics with increased airway reactivity (direct correlation).

**Chemotactic Factors.** Two tetrapeptides, Val/Ala-Gly-Ser-Glu, have been identified as the active substances in the human eosinophil chemotactic factor of anaphylaxis (ECF-A) (16). ECF-A is selectively chemotactic for eosinophils but is also capable of deactivating eosinophils to subsequential attraction by ECF-A or C5a (17). Neutrophil chemotactic factor (NCF) is a neutral protein (molecular weight 750,000) preferentially chemotactic for neutrophils and will deactivate both neutrophils and eosinophils (2). Partially purified NCF provoked by specific antigen and released during exercise in atopic asthmatics promotes both chemokinesis and chemotaxis (18).

**Lysosomal Enzymes.** Chymotrypsin and trypsin are preformed granule-associated proteolytic enzymes found in human mast cells.

Rat and human mast cells have been shown to produce superoxide ($O_{2-}$) during degranulation that is toxic for biologic membranes (19). The enzyme superoxide dismutase present in rat mast-cell granules facilitates the dismutation of $O_{2-}$ to $H_2O_2$ and thus may have a protective effect against this radical toxicity. A rat mast-cell peroxidase tightly bound to the granule matrix has also been described (20), as has a tryptic protease (tryptase) (21). Tryptase has been found in plasma of patients with systemic mastocytosis and after anaphylaxis, especially bee-sting-induced anaphylaxis (22).

**Heparin.** In the rat mast cell, heparin makes up about one-third of the granule by dry weight (23). The rat mast-cell heparin has an average molecular weight of 650,000 and contains a protein core to which glycosaminoglycan side chains are attached. Heparin has many activities, including antithrombin, anticoagulative, and anticomplement action, and the ability to abolish the capacity of eosinophil major basic protein to kill parasites (2).

**Kininogenase.** A kininogenase released from partially purified human lung mast cells in a dose-dependent fashion by anti-IgE has been reported (24). The release of kinin-generating activity parallels that of histamine. Released kinins may contract smooth muscle, increase vascular permeability, and mediate prostaglandin formation.

**Newly Generated Mediators.** Arachidonic acid is the major precursor for the generation of prostaglandins, leukotrienes, thromboxanes, and hydroxy eicosatetraenoic acids (HETE). Products of arachidonic acid metabolism via a cyclooxygenase-dependent pathway constitute the vast majority of prostaglandins and thromboxanes (TX), whereas those via a lipoxygenase pathway constitute the leukotrienes and HETE. The biologically active moiety previously referred to as slow-reacting substance of anaphylaxis (SRS-A) is now known to be composed of sulfidopeptide leukotrienes (LT) $C_4$, $D_4$, and $E_4$. Some of the physiologic effects of the arachidonic acid metabolites are listed in Table 2. It is readily apparent that these compounds have a wide range of activities and are important in the pathogenesis of many allergic diseases such as asthma.

Recent exciting investigations have demonstrated that the mast cell can also generate and release cytokines that are important in several pulmonary reactions (25). These are interleukin-1 (IL-1) and tumor necrosis factor (TNF), which activate endothelial cells, and IL-3, IL-4, IL-5 and granulocyte-monocyte-colony stimulating factor (GM-CSF) (26), which are cytokines that can prime and aid in the recruitment of granulocytes. The secretion of these cytokines is stimulated by antigen after a lag period of hours (27). Therefore, the mast cell may also participate in allergic late-phase reactions and in the recruitment of cells to asthmatic airways by the production of cytokines. Although mast cells influence the pathophysiology of allergic disease such as asthma, allergic rhinitis, and urticaria, mast cells can also regulate bone formation, help repair connective tissue, promote wound healing, and contribute to the integrity of the endothelium.

TABLE 2

*Pathobiologic Events in Asthma Attributable to Actions of Chemical Mediators of Anaphylaxis*

| | |
|---|---|
| Bronchospasm | Histamine ($H_1$ response)<br>SRS-A ($LTC_4$, $LTD_4$, $LTE_4$)<br>Prostaglandins and thromboxane $A_2$<br>Acetylcholine<br>Bradykinin |
| Mucosal edema | Histamine ($H_1$ response)<br>PGE<br>SRS-A ($LTC_4$, $LTD_4$, $LTE_4$) |
| Eosinophil infiltration | Histamine<br>ECF-A<br>HETEs<br>$LTB_4$ |
| Neutrophil infiltration | NCF-A<br>Inflammatory factors of anaphylaxis (IF-A)<br>$LTB_4$<br>HETEs<br>HHT |
| Mucus secretion | Histamine<br>Acetylcholine<br>Alpha-adrenergic agonists<br>Prostaglandins<br>HETEs<br>SRS-A ($LTC_4$, $LTD_4$, $LTE_4$)<br>Macrophage mucus secretagogue (MMS)<br>Prostaglandin generating factor of anaphylaxis (PGF-A) |
| Desquamation | $O_2$, $H_2O_2$, OH<br>Proteolytic enzymes |
| Basement membrane thickening | Proteolytic enzymes |

## Central Role of Mast Cells in Allergic Asthma

The role of mast cells and mast-cell mediators in the pathogenesis of asthma is important. Ample evidence supports this concept, including:

- The effectiveness of cromolyn sodium as both an inhibitor of mast cell secretion and a therapeutic agent in the treatment of asthma.
- The release of neutrophil chemotactic factor of anaphylaxis (NCF) into the circulation of patients with bronchial asthma after inhalation of specific antigen (28–30).
- The release histamine and NCF in exercise-induced asthma (18, 31–33) but not hyperventilation-induced asthma (32, 34–35).
- The release of histamine in late asthmatic reactions.
- The release of NCF (36) in late asthmatic reactions after antigen inhalation in sensitized individuals (37).

Electron microscopic analysis indicates that mast cells comprise 2% of alveolar tissue (4). Thus, the effector cell necessary for allergic reactions to occur in the respiratory tract is present in large numbers and in anatomically strategic locations, such that the release of the mediators of anaphylaxis might account for the pathologic changes noted in asthmatics.

Bronchial asthma is a disease of hyperreactive airways manifested by reversible airway obstruction. The physiologic alterations associated with asthma include increased airway resistance, decreased expiratory flow, decreased vital capacity, increased lung volume, and ventilation-perfusion imbalance. The anatomic alterations in asthma involve all layers of the airway wall and result in the following pathologic changes: bronchospasm, mucosal edema, cellular infiltrates, mucus secretion, desquamation of surface epithelial cells, basement membrane thickening, and goblet cell hyperplasia. Many of the mast-cell mediators can play an important role in each of these pathophysiologic changes and are discussed here in the context of human asthma (Table 2).

## Bronchospasm

The following mast-cell-derived mediators are capable of causing bronchial smooth muscle constriction: histamine; bradykinin; slow-reacting substances of anaphylaxis (SRS-A), which include leukotrienes $C_4$, $D_4$, and $E_4$ ($LTC_4$, $LTD_4$, and $LTE_4$); prostaglandins $PGG_2$, $PGH_2$, and $PGD_2$; and thromboxane $A_2$ ($TxA_2$).

**Histamine.** There are two cellular receptors for histamine, designated $H_1$ and $H_2$. Histamine induces airway obstruction through stimulation of $H_1$ receptors on muscle fibers. Analysis of histamine-induced constriction in vivo reveals that histamine acts directly on muscles (38) and may have a vagally mediated reflex parasympathetic action as well (39). Histamine also dilates small radicles of the pulmonary vascular tree through an $H_1$ receptor and increases the distance between endothelial cells of the venules, thereby increasing the potential for transudation of plasma and for extravasation of leukocytes. The $H_1$ specificity of this response is confirmed by the ability to block this response by specific $H_1$ antagonists (40); $H_2$ receptor blockade has little influence on airway constriction.

Why is it then that $H_1$ antihistamines are clinically not very effective in treating asthma? It has been estimated that the local concentrations of histamine in the lung during in vitro passive anaphylaxis is >10 μm. In vitro studies suggest that nearly equivalent concentrations of $H_1$ antagonists are required for prevention of histamine-induced effects (41). It is unlikely that systemically administered antihistamines are able to achieve these levels (1 to 10 μm). Thus, the lack of antihistamine effectiveness in asthma may in part reflect insufficient pharmacologic blockade rather than an indirect reflection of histamine's actions (see below).

**Prostaglandins.** Products of arachidonic acid metabolism via a cyclooxygenase-dependent pathway constitute the vast majority of prostaglandins in humans. During the course of human lung anaphylaxis, different prostaglandins are generated (41). The appearance of histamine in lung supernatants after immunologic challenge is closely paralleled by the generation of $PGF_{2\alpha}$ and PGE. $PGF_{2\alpha}$, $PGE_1$, $PGD_2$, and thromboxane $B_2$ cause bronchospasm, whereas $PGE_2$ and $PGI_2$ are bronchondilators. It is conceivable that a balance between these opposing influences may affect bronchial tone and play a role in allergic bronchospasm.

There are a number of sources of prostaglandin generation accompanying allergic responses. Histamine, through stimulation of $H_1$ receptors, is responsible for about 50% of the prostaglandins generated during anaphylaxis (41). Histamine-induced prostaglandin generation has also been attributed to muscle contraction. Airway muscle contraction selectively causes PGE generation irrespective of the cause of the muscle contraction, and histamine-stimulated peripheral lung responds with the generation of both $PGF_{2\alpha}$ and PGE. The parenchymal response is specific for histamine and is not related to muscle contraction (42). Mediators other than histamine, such as bradykinin and SRS-A, have also been shown to be capable of causing prostaglandin formation. In the human, products of arachidonic acid metabolism in a cyclooxygenase-dependent pathway constitute the vast majority of PGS.

**Arachidonate Lipoxygenase Metabolites.** The biologically active mediator previously referred to as slow-reacting substance of anaphylaxis (SRS-A) is now known to be composed of leukotrienes (LT) $C_4$, $D_4$, and $E_4$. Leukotrienes are a series of closely related conjugated trienes derived from arachidonic acid through the lipoxygenase pathway. In vivo experiments comparing the effects of SRS-A on airway resistance and lung compliance, and comparisons of the effects of $LTC_4$ and $LTD_4$ on tracheal and parenchymal strips, indicate that these agents are more active in contracting small airways and peripheral lung tissue than larger airways (43). Studies on human bronchial muscle indicate that $LTC_4$ and $LTD_4$ are 1000 times more potent than histamine and 500 times more potent than $PGF_{2\alpha}$ on a molar basis in their ability to cause contraction (44). In the sensitized lung, cyclooxygenase inhibitors (aspirin and indomethacin) may enhance allergic tracheal contraction by diverting arachidonic acid through the lipoxygenase pathway toward the production of SRS-A or related bronchoconstrictive leukotrienes (45, 46). Experiments in normal humans have demonstrated dose-dependent and selective decrements in flow rates at 30% of vital capacity after inhalation of $LTC_4$ or $LTD_4$. $LTC_4$ and $LTD_4$ were 3800 times more potent on a molar basis than histamine in this action.

These studies indicate that in normal humans $LTC_4$ and $LTD_4$ are the most potent bronchoconstrictive substances yet described and that their prolonged duration of effects (15 to 20 minutes) is consistent with a possible role in the mediation of IgE-dependent bronchonconstriction (47).

### Mucosal Edema

Edema of airway mucosa results from an increased capillary permeability with leakage of serum proteins into interstitial areas. Histamine, PGE, bradykinin, $LTC_4$, $LTD_4$, and $LTE_4$ are all capable of causing increased capillary permeability. Complete pharmacologic prevention of allergic-induced vascular permeability would require inhibition of both histamine receptors and arachidonic acid metabolism via the cyclooxygenase and lipoxygenase pathways to inhibit the formation of PGs and LTs respectively.

### Cellular Infiltrates

The mucosa of patients who die in status asthmaticus contains mixed cellular infiltrates consisting of eosinophils, neutrophils, macrophages, monocytes, and plasma cells (48). There are several mediators capable of attracting migrating cells to sites of mast-cell degranulation. Mast-cell degranulation leads to the rapid appearance of histamine, eosinophil chemotactic factors (ECF), and arachidonic acid metabolites such as $LTB_4$ (49), which may initiate the characteristic eosinophil and neutrophil infiltration prominent in the histopathology of status asthmaticus. Also, the mast-cell granule matrix may be a source of mediators that can elicit cellular infiltrates appearing up to 24 hours after the initial antigen-induced mast-cell degranulation event. These delayed cellular infiltrations may also contribute in some way to the airway obstruction seen 4 to 8 hours or 24 hours after antigen bronchial provocation testing.

### Mucus Secretion

Pathologic examinations of status asthmaticus patients who have died almost always reveal diffuse secretions of mucus that appear to contribute significantly to the airways obstruction. Allergic pulmonary reactions in vivo lead to airway constriction and mucus secretion, while in vitro lung anaphylaxis leads to mediator release and increased mucous glycoprotein secretion from cultured airways (50). The association between allergic pulmonary reactions and excessive mucus release suggests that immediate hypersensitivity reactions may lead to the release or generation of secretagogues of mucus.

To identify factors capable of influencing mucus secretion, a model was developed (50–52) employing human airways cultured in the presence of radiolabeled amino sugars. The labeled glycoproteins (mucins) synthesized in vitro may be quantified and potential modulating influences examined. Using human airways in culture (50–52), the role of the mediators of anaphylaxis in mucus release has been studied. Histamine in micromolar concentrations increases mucus release predominantly through $H_2$ receptor stimulation (50). $PGA_2$, $PGD_2$, $PGI_2$, $PGE_1$, and $PGF_{2\alpha}$ at 1 to 100 μm concentrations significantly increased mucous glycoprotein release, whereas $PGE_2$ at 0.01 to 1 μM significantly reduced release (51). Further studies indicate that the monohydroxyeicosatetraenoic acids (HETEs) at nanomolar concentrations increase mucus release (51). Because SRS-A is generated during allergic reactions of human lung tissue at 500 to 5000 units/g lung (53) and has been found to be composed of mixtures of $LTC_4$ and $LTD_4$, the ability of these compounds to release mucus was tested. Both $LTC_4$ and $LTD_4$ produced dose-related increases in mucus production at concentrations of pg/mL (54).

Thus, many mast-cell mediators are potent mucus secretagogues and may possibly contribute to the mucus secretion observed during allergic reactions in vitro and in vivo.

The primary sources of mucus in the human airways are the submucosal glands and, to a lesser extent, the goblet cells. Parasympathetic fibers innervate submucous glands and stimulation causes both smooth muscle contraction and glandular secretion. Asthma is associated with hyperplasia of goblet cells that appear devoid of neural innervation, and the regulation of their secretion is unknown. Both cholinergic and alpha adrenergic neurohormones probably have no effect (50).

A low-molecular-weight mediator generated by human pulmonary alveolar macrophages and human peripheral blood monocytes (55, 56) during phagocytosis and receptor activation has been identified. It is also a mucus secretagogue (55, 56). This molecule, termed macrophage (monocyte) mucus secretagogue (MMS), 2000 daltons in molecule weight, is not a lipid or a prostaglandin but a peptide with approximately 20 amino acids that is resistant to heat inactivation and recurrent freeze and thaw manipulations, but is destroyed by proteases.

MMS is involved in the mucus secretion accompanying inflammatory lung diseases such as pneumonia and acute and chronic bronchitis and in heavy smokers without bronchitis. It is a new mediator of inflammation that regulates mucus secretion. Better understanding of the up-and-down regulation of this mediator may help in the treatment of lung diseases associated with mucus hypersecretion such as asthma and chronic bronchitis.

Mast cells can also release the mediator platelet-activating factor (PAF) when they are stimulated during immediate hypersensitivity reactions. PAF is derived from membrane phosphatidyl-choline. The release of PAF in the airways may cause an increase in mucus release, both by activation of the mucus-producing cells and by causing inflammation in the airways. PAF is a strong stimulator of mucus secretion, and the data from studies in humans (57) suggest that the PAF response is also mediated by activation of the endogenous lipoxygenase pathway in the airways.

## Therapeutic Approach to Allergic Asthma

**Central Role of Mast Cell.** Following a specific allergic reaction, mediators released from mast cells found free in the bronchial lumen or in the epithelial surface loosen the interepithelial cell-tight junctions, allowing the entrance of more allergen to the deeper mast cells (58). The primary and secondary mediators thereby generated induce a further increase in vascular permeability that leads to the entrance of plasma proteins and platelets. The other immediate responses induced by mediator release are smooth muscle constriction, mucus secretion, and leukocyte chemoattraction. Vagal afferent and reflex efferent stimulation are induced by histamine and probably other mediators, which might contribute both to the bronchospasm and to mucous gland secretion. Subacute responses include increased cellular infiltrates, mucosal edema, desquamation, basement membrane thickening, goblet cell hyperplasia, and mucus secretion. These responses may occur because of the continued release of primary and secondary mediators as well as effects caused by the mast-cell granule matrix-derived factors. Thus, many of the pathologic features of asthma may be attributed to mast cell degranulation.

**Controlling Inflammation and Bronchospasm.** All of the above emphasizes that asthma is an inflammatory process of which bronchospasm is only one component. The therapeutic approach must aim at the control of inflammation as well as bronchospasm. The β-adrenergic agents and aminophylline are the most commonly employed bronchodilators, but both can also down-regulate the inflammatory process in some unexplained way. The more specific drugs, corticosteroids and cromolyn, are more clearly in the anti-inflammatory category. It has been reported (59) that albuterol and fenoterol (β-adrenergic agents) were more potent inhibitors of both histamine and neutrophil chemotactic factor release than cromolyn, and that ipratropium (a synthetic atropinelike drug), while equally effective as a bronchodilator, did not inhibit mediator release at all (60). Cromolyn is an excellent and safe anti-inflammatory drug in asthma (61, 62).

**Antihistamines.** Since histamine is an important mediator of allergy, understandably antihistamines are often considered as possible drugs for patients with allergic asthma. Two major types of histamine receptors, $H_1$ and $H_2$, have been identified by using their agonists and antagonists. $H_1$-receptor antagonists are reversible, competitive inhibitors of the actions of histamine on $H_1$ receptors. They decrease vascular permeability, decrease pruritus, and relax smooth muscle. $H_2$-receptor antagonists, such as cimetidine, ranitidine, and famotidine, decrease gastric acid secretions and decrease mucous glycoprotein secretions from human lung airways in vitro (50). In some well-designed, double-blind studies, $H_1$-receptor antagonists, such as chlorpheniramine, hydroxyzine, clemastine, terfenadine, and cetirizine have exhibited a modest, dose-related bronchodilator effect. In bronchial challenge studies, pretreatment with an $H_1$-receptor antagonist prevents histamine, antigen, hyperventilation and exercise-induced asthma (63–65), but not methacholine-induced asthma (66). Long-term bronchodilator effects of $H_1$-receptor antagonists, such as chlorpheniramine or astemizole, administered for several months to patients with chronic asthma, have been demonstrated (67, 68). Although $H_1$-receptor antagonists are not drugs of first choice for acute or chronic asthma, previous concerns about the potential adverse effects of $H_1$-receptor antagonists in asthma have been exaggerated. If $H_1$-receptor antagonists are required for other allergic manifestations in a patient with asthma, they should not be withheld.

**Drugs to Block Platelet-Activating Factor.** Although the bronchial hyperreactivity in asthmatics is clearly associated with infiltration of inflammatory cells, there is also an inflammation-independent component to bronchial hyperreactivity. Some current investigators in asthma physiology have implicated the SRS-A (leukotrienes C4, D4, and E4) as responsible for this phenomenon (69). Platelet-activating factor (PAF), derived from membrane lipid by a reaction involving phospholipase A, together with a 5-lipoxygenase enzyme, is also involved in the generation of the leukotrienes C4, D4, and E4 from membrane lipid. A wide range of inflammatory cells types (including the mast cell) can generate and release significant quantities of PAF in response to either allergic or nonallergic stimuli (70). PAF can affect the tissue either indirectly, by inducing platelet aggregation with the release of mediators, or directly, as by platelet-independent endothelial injury. If PAF is inhaled, it causes immediate bronchonconstriction, and also a prolonged increase in nonspecific bronchial responsiveness in nonasthmatic subjects (71). When injected intravenously, it causes pulmonary vasoconstriction, systemic vasodilation, microthrombi in the capillaries, and platelet-independent injury to vascular endothelium. It is a potent activator and chemoattractant for polymorphonuclear leukocytes and eosinophils, as well as an aggregator of platelets. It is known to increase mucus secretion in rodents' airways in vitro (72) as well as from human airways in culture (57). It is therefore possible that PAF, which was first described as a product of basophils following allergen activation, is an important mediator in asthma.

PAF has received great attention for its role in asthma as well as in endotoxin shock. PAF and leukotrienes are among the most important substances that regulate inflammatory processes in asthma. In an asthmatic reaction, PAF binds to eosinophils, causing them to release cytoplasmic granules containing proteinases and histamines, leading to inflammation and bronchospasm. PAF can also amplify the reaction by inducing other types of white blood cells to produce more PAF and other lipid mediators that further contribute to inflammation. Researchers from Japan reported recently that they have cloned and sequenced the gene that codes for the PAF receptor (73). This new work, which is based on studies of guinea pig lung, may lead to new drugs to block the binding of PAF to its receptor and lead to new treatment of asthma and endotoxin shock.

**Corticosteriods.** The corticosteroids (cortisone) are the most successful anti-inflammatory agents in the treatment of asthma. The exact mechanism of their action is still not fully understood, but it is thought that they act by inducing the synthesis of a protein (lipomodulin) that is a specific inhibitor of phospholipase $A_2$, the first enzyme involved in the synthesis of both PGS and leukotrienes from membrane arachidonic acid. Various glucocorticoids inhibit baseline mucus secretion and stimulated mucus secretion from human airways (74). The effect has a latent period of 8 to 16 hours indicating a possible need for the generation of another mediator by these drugs. It is presumed that the glucocortoid effect on the airways mucus release is mediated through lipocortin production and subsequent inhibition of arachidonic acid release from the airways (75–76). Steroids are so far the most effective drugs in the treatment of mucus hypersecretion associated with airway inflammation.

**Other Pharmacologic Agents.** With this new and growing understanding of asthma pathophysiology, there is a greatly widened spectrum of pharmacologic attacks available against the cascade of events causing asthma. In addition to PAF receptor blockers, investigators are trying to find a more specific inhibitor of phospholipase A (like cortisone), there is a search for clinically useful inhibitors of the 5-lipoxygenase pathway, and active research to find and synthesize competitive inhibitors of the leukotrienes. The unique role of the enzyme 5-lipoxygenase in the production of leukotrienes makes it an attractive target. Two studies have focused on an orally active agent capable of inhibiting 5-lipoxygenase in airway obstruction induced by cold air and antigen-induced nasal congestion (77, 78). In another investigation the effect of an orally administered leukotriene $D_4$-receptor antagonist on exercise-induced bronchoconstriction has been reported (79). The short-term successful experimental treatments reported in these studies offer encouraging new therapeutic modalities for the control of common allergic disorders.

Finally, mention should be made of ketotifen (Zaditen), a new agent well known and used in

Europe but still not available for use in the United States. The antiallergic effects of this drug were initially characterized by its ability to prevent passive cutaneous anaphylaxis in the rat. It was then found to be a potent $H_1$-antihistamine. It will inhibit the release of SRS-A in some systems, and also reverses adrenergic tachyphylaxis (steroidlike effect). It has been reported that ketotifen can block the effects of PAF in some systems (80). There is still a controversy about its efficacy in asthma, but it may be the first in another powerful class of drugs with anti-inflammatory capabilities in the treatment of asthma.

## Summary

The role of mast cells in human biology is important and complex. They probably play a dual role both in protecting the host and in causing disease symptoms. For example: mast-cell-mediated vasopermeability is an important manifestation of asthma, but it also provides access of antibody and other blood factors which are brought to the tissue and help in defense. Therefore, these cells, with their abundant mediators, may play an important role in both homeostasis and disease.

## References

1. Riley JF. The mast cells. Edinburgh: E&S Livingston Ltd., 1959; 5–17.
2. Metcalfe DD, Kaliner M, Donlon MA. The mast cell. CRC Critical Rev Immunol 1981; 3:23–74.
3. Wasserman SI. The lung mast cell: its physiology and potential relevance to defense of the lung. Environmental Health Perspect 1980;35:153–164.
4. Fox B, Bull TB, Guz A. Mast cells in the human alveolar wall: an electromicroscopic study. J Clin Pathol 1981; 34:1333–1342.
5. Metzger H, Goetz A, Kanellopoulos J, Holowka D, Fewtrell C. Structure of the high-affinity mast cell receptor for IgE. Fed Proc 1982; 41:8–11.
6. Kulczycki A, McKearney TA, Parker CW. The rat basophilic leukemia cell receptor for IgE: I. Characterization as a glycoprotein. J Immunol 1976; 117:661–665.
7. Kaliner M. Human lung and anaphylaxis: evidence that cyclic nucleotides modulate the immunologic release of mediators through effects on microtubule assembly. J Clin Invest 1977; 60:951–959.
8. MacGlashan DW, Schleimer RP, Peters SP, et al. Generation of leukotrienes by purified human lung mast cells. J Clin Invest 1982; 70:747–751.
9. Röhlich P, Anderson P, Uvnas B. Mast cell degranulation as a process of sequential exocytosis. Acta Biol Acad Sci Hung 1971; 22:197–202.
10. Sullivan TJ. Diacylglycerol metabolism and the release of mediators from mast cells. In: Becker EL, Austen KF, eds. Biochemistry of the acute allergic reactions. Fourth International Symposium. New York: Alan R. Liss 1981; 229–238.
11. Kaliner MA. Mast cell-derived mediators and bronchial asthma. In: Hargreave FE, ed. Airway reactivity. Mississauga, Ontario, Canada: Astra Pharmaceuticals Canada Ltd., 1980; 175–187.
12. Denberg JA. Phylogeny and ontogeny of basophils, mast cells and eosinophils. In: Holgate S, ed. Mast cells, mediators and disease. Dordrecht, The Netherlands: Kluwer Academic, 1988; 1–29.
13. Brodide DH, Wasserman SI, Alvaro-Garcia J, Zvaifler N, Firestein G. Transforming growth factor beta selectively inhibits IL-3 dependent mast cell proliferation without affecting mast cell function or differentiation. J Immunol 1989; 143:1591–1597.
14. Baeza ML, Haak-Frendscho M, Satnick S, Kaplan AP. Responsiveness to human mononuclear cell-derived histamine-release factor: studies of allergic status and the role of IgE. J Immunol 1988; 141:2688–92.
15. Casale TB, Kaliner M. Allergic reactions in the respiratory tract. In: Bienenstock J, ed. Immunology of the lung. New York: McGraw-Hill, 1984.
16. Goetzl EJ, Austen KF. Purification and synthesis of eosinophilotactic tetrapeptides of human lung tissue: identification as eosinophil chemotactic factor of anaphylaxis. Proc Natl Acad Sci USA 1975; 72:4123–4127.
17. Wasserman SI, Whitmer D, Goetzl EJ, Austen KF. Chemotactic deactivation of human eosinophils by the eosinophil chemotactic factor of anaphylaxis. Proc Soc Exp Biol Med 1975; 148:301–306.
18. Lee TH, Nagy L, Nagakura T, Walport MJ, Kay AB. Identification and partial characterization of an exercise-induced neutrophil chemotactic factor in bronchial asthma. J Clin Invest 1982; 69:889–899.
19. Henderson WR, Kaliner M. Immunologic and non-immunologic generation of superoxide from mast cells and basophils. J Clin Invest 1978; 61:187–196.
20. Henderson WR, Kaliner M. Mast cell granule peroxidase: location secretion and SRS-A inactivation. J Immunol 1979; 122:1322–1328.
21. Schwartz LB, Austen KF. The mast cell and mediators of immediate hypersensitivity. In: Samter M, Talmage DW, Frank MM, Claman HN, eds. Immunological diseases. Boston: Little Brown, 1988; 157–201.
22. Schwartz LB. Tryptase, a mediator of human mast cells. Allerg Clin Immunol 1990; 86:594–598.
23. Lagunoff D, Pritzl P. Characterization of rat mast cell granule proteins. Arch Biochem Biophys 1976; 173:554–563.
24. Proud D, Schulman ES, MacGlashan DW, Pierce JV, Newball HH. Anaphylactic release of a kininogenase from purified human lung mast cells. Clin Res 1982; 30:165 (abstract).
25. Plaut M, Pierce JH, Watson CJ, Hanley-Hyde J, Nordan RP, Paul WE. Mast cell lines produce lymphokines in response to cross-linkage of $Fc_\epsilon RI$ or to calcium ionophores. Nature 1989; 339:64–67.
26. Wodnar-Filipowicz A, Heusser CH, Maroni C. Production of the haemopoietic growth factors GM-CSF and interleukin-3 by mast cells in response to IgE receptor-mediated activation. Nature 1989; 339:150–2.
27. Burd PR, Rogers HW, Gordon JR, et al. Interleukin 3-dependent and independent mast cells stimulated with IgE and antigen express multiple cytokines. J Exp Med 1989; 170:245–57.
28. Atkins PC, Norman M, Weiner M, Zweiman B. Release of neutrophil chemotactic activity during immediate hypersensitivity reactions in humans. Ann Intern Med 1977; 86:415–418.
29. Atkins PC, Norman M, Zweiman B. Antigen-induced che-

motactic activity in man: correlation with bronchospasm and inhibition by disodium cromoglycate. J Allerg Clin Immunol 1978; 62:149–155.

30. Atkins PC, Norman M, Zweiman B, Rosenblum F. Further characterization and biologic activity of ragweed antigen-induced neutrophil chemotactic activity in man. J Allerg Clin Immunol 1979;64:251–258.
31. Lee TH, Brown MK, Nagy L, Causon R, Walport MJ, Kay AB. Exercise-induced release of histamine and neutrophil chemotactic factor in atropic asthmatics. J Allerg Clin Immunol 1982; 70:73–81.
32. Barnes PJ, Brown MJ. Venous plasma histamine in exercise and hyperventilation-induced asthma in man. Clin Sci 1981; 61:159–162.
33. Anderson SD, Bye PTP, Schaeffel RE, Seale JP, Taylor KM, Ferris L. Arterial plasma histamine levels at rest, during and after exercise in patients with asthma: effects of terbutaline aerosol. Thorax 1981; 36:259–267.
34. Deal EC Jr, Wasserman SI, Soter NA, Ingram RH Jr, McFadden ER Jr. Evaluation of role played by mediators of immediate hypersensitivity in exercise-induced asthma. J Clin Invest 1980; 49:659–665.
35. Nagakura T, Lee TH, Asoufi B, Denison DM, Newman Taylor AJ, Kay AB. Neutrophil chemotactic factor (NCF) in exercise and hyperventilation-induced asthma (abstr). Thorax 1983.
36. Kay AB, Lee TH. Neutrophil chemotactic factor of anaphylaxis. J Allerg Clin Immunol 1982; 70:317–320.
37. Nagy L, Lee Th, Kay AB. Neutrophil chemotactic activity in antigen-induced late asthmatic reactions. N Engl J Med 1982; 306:497–501.
38. Rosenthal RR, Norman PS, Summer WR, Permutt S. Role of the parasympathetic system in antigen-induced bronchospasm. J Appl Physiol 1977; 42:600–606.
39. Yu DYC, Galant SP, Gold WM. Inhibition of antigen-induced bronchospasm by atropine in asthmatic patients. J Appl Physiol 1972; 32:823–828.
40. Casterline CL, Evans R. Further studies on the mechanism of human histamine-induced asthma. J Allerg Clin Immunol 1977; 59:420–425.
41. Platshon L, Kaliner M. The effect of the immunologic release of histamine upon human lung cyclic nucleotide levels and prostaglandin generation. J Clin Invest 1978; 62:1113–1121.
42. Steel L, Platshon L, Kaliner M. Prostaglandin generation by human and guinea pig lung tissue: comparison of parenchymal and airway responses. J Allerg Clin Immunol 1979; 64:287–293.
43. Spannhake EW, Hyman AL, Kadowitz PJ. Bronchoactive metabolites of arachidonic acid and their role in airway function. Prostaglandins 1981; 22:1013–1026.
44. Dahlen SE, Hedqvist P, Hammarstrom S, Samuelson B. Leukotrienes are potent constrictors of human bronchi. Nature 1980; 288:484–486.
45. Burka JF, Paterson NAM: Evidence for lipoxygenase pathway involvement in allergic tracheal contraction. Prostaglandins 1980; 19:499–515.
46. Paterson NAM, Burka JF, Craig ID. Release of slow reacting substance of anaphylaxis from dispersed pig lung cells: effect of cyclooxygenase and lipoxygenase inhibitors. J Allerg Clin Immunol 1981; 67:426–434.
47. Weiss JW, Drazen JM, McFadden ER, et al. Comparative bronchoconstriction effects of histamine and leukotrienes C and D (LTC and LTD) in normal human volunteers. Clin Res 1982; 30:517 (abstract).
48. Kaliner M, Blennerhassett J, Austen KF. Bronchial asthma. In: Meischer PA, Muller-Eberhard HJ, eds. Textbook of immunopathology. New York: Grune and Stratton, 1976; 387–402.
49. Lewis RA, Goetzl EJ, Drazen JM, Soter NA, Austen KF, Corey EJ. Functional characterization of synthetic leukotriene B and its stero-chemical isomer. J Exp Med 1981; 154:1243–1248.
50. Shelhamer JH, Marom Z, Kaliner M. Immunologic and neuropharmacologic stimulation of mucous glycoprotein release from human airways in vitro. J Clin Invest 1980; 66:1400–1408.
51. Marom Z, Shelhamer JH, Kaliner M. The effects of arachidonic acid, monohydroxyeicosatetraenoic acid and prostaglandins on the release of mucous glycoprotein from human airways in vitro. J Clin Invest 1981; 67:1695–1702.
52. Kaliner M, Shelhamer J, Borson B, Nadel J, Patow C, Marom Z. Human respiratory mucus. Am Rev Respir Dis 1986; 134:612–621.
53. Kaliner M, Wasserman SI, Austen KF. Immunologic release of chemical mediators from human nasal polyps. N Engl J Med 1973; 289:277–281.
54. Marom Z, Shelhamer JH, Bach MK, Morton DR, Kaliner M. Slow-reacting substances, leukotrienes C4 and D4 increase the release of mucus from human airways in vitro. Am Rev Respir Dis 1982; 126:449–451.
55. Marom Z, Shelhamer JH, Kaliner M. Human pulmonary macrophage-derived mucus secretagogue. J Exp Med 1984; 159:844–860.
56. Marom Z, Shelhamer JH, Kaliner M. Human monocyte-derived mucus secretagogue. J Clin Invest 1985; 75:191–198.
57. Goswami S, Ohashi M, Stathas P, Marom Z. Platelet activating factor (PAF) stimulates secretion of respiratory glycoconjugates from human airways in culture. J Allerg Clin Immunol 1989; 84:726–734.
58. Hogg JC, Pare PD, Boucher R, Michond MC, Guerzon G, Moroz L. Pathologic abnormalities in asthma. In: Lichtenstein LM, Austen KF, eds. Physiology, immunopharmacology and treatment (Second International Symposium). New York: Academic Press, 1977; 1–14.
59. Howarth PH, Durham SR, Lee TH, et al. Influence of albuterol, cromolyn sodium and ipratropium bromide on the airway and circulating mediator responses to allergen bronchial provocation in asthma. Am Rev Respir Dis 1985; 132:986–992.
60. Greenspon LW, Morrissey WL. Factors that contribute to inhibition of methacholine-induced bronchoconstriction. Am Rev Respir Dis 1986; 133:735–739.
61. Reed CE. New therapeutic approaches in asthma. J Allerg Clin Immunol 1986; 77:537–543.
62. Bernstein IL. Cromolyn in the treatment of asthma; coming of age in the United States. J Allerg Clin Immunol 1985; 76:381–388.
63. Rafferty P, Holgate ST. Terfenadine (seldane) is a potent and selective histamine $H_1$ receptor antagonist in asthmatic airways. Am Rev Respir Dis 1987; 135:181–184.
64. Brik A, Tashkin DP, Gong H, Dauphinee B, Lee E. Effect of cetirizine, a new histamine $H_1$ antagonist, on airway dynamics and responsiveness to inhaled histamine in mild asthma. J Allerg Clin Immunol 1987; 80:51–56.
65. Badir M, Beaumont D, Orehek J. Attenuation of hyperventilation-induced bronchospasm by terfenadine: a new antihistamine. J Allerg Clin Immunol 1988; 81:437–440.
66. Patel KR. Effect of terfenadine on methacholine-induced bronchoconstriction in asthma. J Allerg Clin Immunol 1987; 79:355–358.

67. Lewiston NJ, Johnson S, Sloan E. Effect of antihistamine on pulmonary function of children with asthma. J Pediatr 1982; 101:458–460.
68. Holgate ST, Emanuel MB, Howarth PH. Astemizole and other $H_1$-antihistamine drug treatment of asthma. J Allerg Clin Immunol 1985; 76:375–380.
69. Adelroth E, Morris MM, Hargreave FE, et al. Airway responsiveness to leukotrienes C4 and D4 and to methacholine in patients with asthma and normal controls. N Engl J Med 1986; 315:480–484.
70. Morley J, Page CP, Sanjar S. Pulmonary responses to platelet-activating factor. Prog Resp Res 1985; 19:117–125.
71. Cuss FM, Dixon CMS, Barnes PJ. Effects of inhaled platelet activating factor on pulmonary function and bronchial responsiveness in man. Lancet 1986; 2:189–196.
72. Adler KB, Schwartz JE, Anderson WH, Welton AF. Platelet activating factor stimulates secretion of mucin by explants of rodent airways in organ culture. Exp Lung Research 1987; 13:25–43.
73. Honda Z, Nakamara M, Miki I, et al. Cloning by functional expression of platelet-activating factor receptor from guinea-pig lung. Nature 1991; 349:342–346.
74. Marom Z, Shelhamer J, Alling D, Kaliner M. The effect of corticosteroids on mucous glycoprotein secretion from human airways in vitro. Am Rev Respir Dis 1984; 129:62–65.
75. Hirata F. Role of lipocortins in cellular function as a second messenger of glucocorticoids. In: Schleimer RP, Claman HN, Aronsky A, eds. Anti-inflammatory steroids: basic and clinical aspects. New York: Academic Press, 1989; 67–95.
76. Lundgren JD, Hirata F, Marom Z, et al. Dexamethasone inhibits respiratory glycoconjugate secretion from feline airways in vitro by the induction of lipocortin (lipomodulin) synthesis. Am Rev Respir Dis 1988; 137:353–357.
77. Israel E, Dermarkarian R, Rosenberg M, Sperling R, Taylor G, Rubin P, Drazen J. The effects of a 5-lipoxygenase inhibitor on asthma induced by cold, dry air. N Engl J Med 1990; 323:1740–1744.
78. Knapp HR. Reduced allergen-induced nasal congestion and leukotriene synthesis with an orally active 5-lipoxygenase inhibitor. N Engl J Med 1990; 323:1745–1748.
79. Manning PJ, Watson RM, Margolskee DJ, Williams VC, Schwartz JI, O'Byrne PM. Inhibition of exercise-induced bronchoconstriction by MK-571, a potent leukotriene $D_4$-receptor antagonist. N Engl J Med 1990; 323:1736–1739.
80. Morley J, Page CP, Mazzoni L, et al. Effects of ketotifen upon responses to platelet activating factor: a basis for asthma prophylaxis. Ann Allergy 1986; 56:335–340.

# Occupational Airway Diseases

E. Neil Schachter, M.D.

> There are many things that a doctor on his first visit to a patient ought to find out either from the patient or from those present. For so runs the oracle of our inspired teacher: "When you come to a patient's house, you should ask him what sort of pains he has, what caused them, how many days he has been ill, whether the bowels are working and what sort of food he eats." So says Hippocrates in his work Affections. I may venture to add one more question: What occupation does he follow? In medical practice however, I find that attention is hardly ever paid to this matter, or if the doctor in attendance knows it without asking, he gives little heed to it, though for effective treatment evidence of this sort has the utmost weight.—B. Ramazzini

The entreaty voiced to the reader in 1713 by Ramazzini (1) remains pertinent to the twentieth-century physician concerned with airway disease.

Airway injury in the occupational setting is frequent but generally underrecognized. Injury may vary in severity from increases in reported respiratory symptom frequency in groups of workers to frank pulmonary edema and even death in individual patients. Most commonly these injuries appear as well-recognized symptom complexes such as asthma or bronchitis. The difficulty facing the physician who suspects an occupational airway disease is to sort out whether the illness is truly related to workplace exposure.

Asthma is a common disease characterized by reversibility, either spontaneously or with therapy. Precise definitions, however, are elusive, since it remains unclear whether the disorder is a single disease with well-defined genetic and pathophysiologic antecedents or the final common pathway of airway injury by many unrelated agents. Most clinicians, nevertheless, now recognize that asthma is distinct from other chronic lung diseases that cause obstruction (emphysema, chronic bronchitis, bronchiectasis) and from nonpulmonary causes of wheezing (for example, congestive heart failure). The association of asthma with an ever-growing list of occupational and environmental agents has made the characterization of these relationships and the study of the mechanisms by which they promote asthma a challenge to clinician and researcher.

Chronic bronchitis is a well-defined clinical syndrome characterized by a history of cough and expectoration for three months yearly for at least two years. The association of chronic bronchitis with the workplace environment is as challenging to diagnose as occupational asthma. In particular, the study of occupational bronchitis may offer insight as to how occupational and other environmental irritants (for example, cigarette smoke) interact to cause chronic lung disease.

The study of occupational airway disease involves not only clinical skills but also an understanding of epidemiologic findings and a familiarity with the work settings involved. Diagnosis and management of these disorders requires that the physician be aware of the unique and varied mechanisms which characterize occupational airway diseases, the environmental and special diagnostic testing available for evaluation, and the therapeutic and preventive options necessary for the management of this unique group of patients.

**Definitions.** Airway diseases associated with occupational exposures result in four unique clinical syndromes (Table 1). Although each of these syndromes is clearly defined and suitable as a diagnosis, a central issue in classification revolves around the question of sensitization.

Industrial bronchitis is defined, as is nonoccupational chronic bronchitis, by chronic cough and expectoration resulting from exposure to irritant gases, fumes, or dusts, usually at relatively low concentrations for long periods. It is a nonimmunologic illness without a latency period and not reflecting specific sensitivity to a single agent. Although cigarette smoking is often a confounding variable in assessing an individual worker with chronic bronchitis (1a), epidemio-

---

Supported in part by grants from the National Institute for Occupational Safety and Health, Centers for Disease Control RO1-OH02593, and the Henry and Catherine Gaisman Fund, New York, NY.

From the Department of Medicine and Community Medicine, Mount Sinai School of Medicine, New York. Address requests for reprints to E. Neil Schachter, M.D., Mount Sinai Medical Center, One Gustave L. Levy Place, Box 1232, New York, NY 10029.

TABLE 1
*Characteristics of Occupational Airway Diseases*

| | Latent period | Sensitization | Airway hyperreactivity |
|---|---|---|---|
| Industrial bronchitis | No | No | +/− |
| Reactive airway dysfunction syndrome (RADS) | No | No | + |
| Bronchiolitis Obliterans | No | No | − |
| Occupational Asthma | Yes | Yes | + |

logic studies from a number of industries indicate the independent effect of these exposures, particularly when contrasting smoking and nonsmoking workers (2).

Reactive airway dysfunction syndrome (RADS) is characterized by chronic airway obstruction and irritability, frequently resulting from a single exposure to high concentrations of gases, vapors, or fumes (3). Despite the relatively short exposure involved, abnormalities, particularly airway hyperreactivity, may persist for years. No sensitization appears to be involved.

Bronchiolitis obliterans is a well-characterized inflammatory response of the terminal airways. Such inflammation results from a usually massive exposure to irritant gases (for example, nitrogen dioxide) which penetrate to the lower airways (4). Characteristically, after initial, relatively mild symptoms of mucous membrane irritation, explosive pulmonary edema follows. If the patient survives, bronchiolitis obliterans may develop. Resolution can occur but frequently there is persistent chronic obstructive lung disease.

Occupational asthma has generally been defined in the narrow sense (5) of airway sensitivity to a specific occupational agent; sensitization usually follows an initial exposure period. Once established, the disease is uniquely and consistently triggered by the sensitizing agent; however, as with nonoccupational asthma, the airway is generally irritable and responds excessively to challenge with nonspecific stimulants. Early removal from the offending agent may abort the disease; however, hyperreactivity once established frequently persists after the worker is removed from the sensitizer.

A less restrictive definition of occupational asthma (6) suggests that it represents airway obstruction, usually reversible, caused by the inhalation of substances or materials that a worker manufactures or uses directly or that are incidentally present at the work site. Such a definition includes both immunologic and nonimmunologic processes. In principle such a definition would include individuals with preexisting asthma whose disease is exacerbated by the workplace.

Occupational airway disease, like its nonoccupational counterpart chronic obstructive pulmonary disease (COPD), does not always fit conveniently into any one of these four categories. Overlap syndromes with elements of bronchitis, reversible bronchospasm, and irreversible chronic airway obstruction may combine in any single individual, and may vary with time.

**Prevalence.** Asthma is a common disease of the airways, affecting up to as many as 10.6% of the general population (7). A gradual increase in the morbidity and mortality of this disease has been noted in this and other countries in the 1980s; it causes an average loss of 15 days per year from work or school. Although chronic asthma is generally considered reversible, evidence from prospective epidemiologic studies is growing that chronic asthma can be associated with progressive lung function impairment (8, 9). Prevalences similar to those of asthma among adults have been documented for the symptoms of chronic bronchitis in population studies (9). Nonspecific respiratory symptoms such as cough and wheeze are common in the general population (10) but have been documented to be 2 to 3 times more frequent in selected dusty industries (11–15).

The prevalence of occupational airway disease is unknown, but some general conclusions are possible. One can expect the prevalence of bronchitic symptoms among workers to rise with the dustiness of the industry. In a recent analysis from the Harvard 6-city study, a significant trend emerged of increasing bronchitis with increasing annual levels of particulate matter less than 15 μm in diameter (16). Korn found increasing prevalences of most respiratory symptoms (chronic cough, phlegm, persistent wheeze, dyspnea, lung function impairment) associated with increasing exposures to gases, fumes, and dust in the workplace (17). More generally: in a recent survey by Schwartz et al. over a 3-month period, nearly 75% of 534 patients in a primary care clinic reported prior occupational exposure to at least one potentially toxic agent, and over 30% claimed exposure to at least four such agents (18).

Estimates of the prevalence of occupationally attributable asthma among new cases of diagnosed asthma have ranged from 2% to 15% (19–21); however, individual agents vary in toxicity from idiosyncratic to very frequent. For example, it has been estimated that some 5% of workers exposed to western red cedar dust develop occupational asthma (22), whereas up to 45% of workers exposed to proteolytic enzymes may develop the disorder (23).

**Etiology and Mechanisms.** Agents causing or associated with occupational airway diseases have been classed as of 3 kinds by physical and chemical characteristics:

- Gases and fumes ($SO_2$, $NO_2$, $NH_3$)
- Large molecular weight substances (glycoproteins, vegetable gums, animal products, plant constituents, insect constituents)
- Low molecular weight substances (acid anhydrides, precious metals, antibiotics, isocyanates)

The clinical syndromes associated with exposure to gases and fumes tend to involve exacerbation of preexisting asthma or RADS. Gases and fumes have well-characterized irritant effects that are usually dose related, nonsensitizing, and more irritating to susceptible individuals such as asthmatics (24). Not all gases with irritating mucous membrane properties are equally dangerous to the airways. For example, formaldehyde, which has far greater potential to irritate the mucous membranes of the eyes and nose than $SO_2$ or $NO_2$, has little or no effect on the airway at the usual occupational levels (25), although sensitization may occur (26). The solubility of the gas and its ability to reach the lower airway is felt to be an important factor in promoting airway effects. Gases such as formaldehyde which are highly soluble and are maximally absorbed in the upper airways may not reach the lower airways in concentrations great enough to cause damage or irritation.

Compounds of low molecular weight known to induce occupational asthma have been arbitrarily defined as having a molecular weight of less than 1000 daltons. These agents are in general simple organic molecules (isocyanates), drugs (antibiotics), or metals (platinum) used in a wide variety of industries (Table 2). Agents extensively studied include platinum salts (27), acid anhydrides (28, 29) such as trimellitic anhydride and tetrachlorophthalic anhydride, isocyanates (30, 31), plicatic acid (32), and antibiotics (33, 34).

Many of these agents are known to act by causing a classic immunologic sensitization by acting as haptens (acid anhydrides, platinum salts); however, for others the immunologic mechanism, although suspected, has not been fully characterized (isocyanates, Western red cedar). Preexisting atopy does not appear to be a risk factor for development of disease with these agents.

Agents of high molecular weight are in general proteins, peptides, or polysaccharides of plant or animal origin (Table 2) which elicit classic humoral sensitivity via IgE or occasionally IgG antibodies. Affected individuals frequently exhibit general signs of atopy such as preexisting allergic rhinitis or eczema. When exposed to these agents either in the natural setting or in a challenge situation, the sensitized individual experiences immediate or biphasic (dual) reactions, the latter characterized by the onset of delayed airway obstruction a number of hours following challenge (Fig. 1).

The mechanism of disease in asthma due to high molecular weight compounds is felt to be identical with asthma due to common allergens such as house dust or pollens. In such patients specific IgE antibodies develop following repeated exposure and lead to an interaction with antigen on the surface of immunoactive cells (for example, mast cells). This in turn leads to the release of both preformed mediators (histamine, eosinophil chemotactic factor) as well as de novo synthesized mediators (prostaglandins, leukotrienes) (35). By contrast, the mechanism involved in airway disease attributable to low molecular weight agents involves inflammation (36, 37), and the role of classic IgE-mediated disease is uncertain, since these antibodies are not invariably associated with disease (38, 39).

Certain host characteristics are also felt to be important in the development of occupational lung disease:

- Atopy
- Cigarette smoking
- Airway hyperreactivity
- Preexisting lung disease
- Genetic factors

The way in which these host characteristics may interact with the workplace include:

- Exacerbation of *preexisting* airway disease by factor(s) in the workplace
- Development of *new* airway disease as a result of factor(s) in the workplace
- Interaction of *preexisting* susceptibility, nonoccupational environmental factor(s) (air pollution, cigarette smoking) and the occupational environmental trigger

TABLE 2

*Characteristics of Low- and High-Molecular-Weight Substances Associated with Occupational Airway Disease*

| | | | Tests | | |
|---|---|---|---|---|---|
| Substance/examples | Industry | Mechanism | Specific skin test | Specific IgE | Bronchial provocation |
| Low molecular weight | | | | | |
| Isocyanates | | ? | +/− | +/− | + |
| Toluene diisocyanate | Plastic<br>Polyurethane | | | | |
| Anhydrides | | Allergic | + | + | + |
| Phthalic anhydride | Plastic, | | | | |
| Trimellitic anhydride | Rubber and resin | | | | |
| Wood dust | | ? | +/− | +/− | + |
| Western red cedar | Saw mills | | | | |
| California redwood | Carpentry<br>Construction | | | | |
| Metals | | Allergic | + | | |
| Platinum | Platinum<br>Refineries | | | | |
| Fluxes | | ? | | | + |
| Colophony | Electronic | | | | |
| Drugs | | Allergic | +/− | | + |
| Antibiotics | Pharmaceuticals | | | | |
| Textile dust | | | | | |
| Cotton<br>Hemp | Textile mills | Nonimmunology release of mediators | − | − | + |
| Large molecular weight | | | | | |
| Animal Proteins | | Allergic | + | + | + |
| Lab animals | Laboratories<br>Veterinarians | | | | |
| Fish/shellfish | Fisheries<br>Crab processing | | | | |
| Birds | Pigeon breeding<br>Poultries | | | | |
| Insects | Laboratories,<br>Entomologist<br>Laboratories | | | | |
| Vegetable dusts | | Allergic | + | + | + |
| Grain dust | Graineries | | | | |
| Flour | Bakeries | | | | |
| Tea/coffee | Food processing | | | | |
| Castor bean | Oil industries | | | | |
| Enzymes | | Allergic | + | + | + |
| Bromelin | Pharmaceutical | | | | |
| Papain | Laboratory | | | | |
| Pepsin | Pharmaceutical | | | | |
| Trypsin | Pharmaceutical | | | | |

**Diagnosis.** The diagnosis of occupational airway disease involves two interrelated approaches: demonstration and characterization of the airway disease (does the worker indeed have airway disease? what is its nature?); and establishment of the work-relatedness of the disease, which is often the most demanding.

The diagnostic approaches and tests used to delineate occupational airway disease include:

- Patient history
- Physical examination
- Chest X-ray
- Pulmonary function studies
- Skin tests
- Serum precipitins
- Nonspecific airway reactivity
- Specific inhalation challenge testing
- Bronchial alveolar lavage
- Workplace studies

Both bronchitis and asthma are clinical diagnoses defined by history and confirmed by lung function studies (for example, reversibility of obstruction by bronchodilator, airway reactivity to methacholine).

A careful occupational history, including review of current and past employment, is essential. This involves an understanding of the nature and variety of workplace agents known or suspected to be airway irritants or sensitizers as well as familiarity with workplace exposures. For each job, the physician should define the nature of the task, its physical setting, agents used, byproducts of the process (dusts from grinding, carding, or sawing; gases given off by chemical reactions), and related processes that occur contiguously. Physical conditions (temperature, for example), level of activity, environmental pollutants (including tobacco smoke) may be important co-factors or even primary causes of disease.

Work-relatedness is frequently suggested by onset of symptoms within the context of the workplace (for example, specialized settings such as the cardroom in byssinosis), absence of previous airway or allergic disease, improvement of symptoms on weekends, vacations, or job transfer, and presence of similar symptoms, whatever their severity, among co-workers. A physician-imposed work restriction with careful follow-up (diaries, peak flow measurements) can be most useful in tracking work-relatedness. In some cases the clinical setting may be pathognomonic, for example the development of Monday dyspnea in textile workers or the sudden onset of respiratory symptoms and laboratory findings of hyperreactivity following heavy exposure to an irritant (e.g. RADS).

In many patients the findings may be confusing—symptoms are more prominent at the end of a workday or at night (Fig. 1), and multiple confounding issues (cigarette smoking, preexisting allergies) are involved. Workers with childhood airway diseases that have subsided in adulthood may pose a difficult diagnostic problem. In workers with established fixed airway disease (chronic bronchitis, chronic airway obstruction) workplace associations may be hard to establish and symptoms may persist through weekends and vacations and even into retirement.

To resolve such puzzles, knowledge of the natural history of disease for a specific agent established in general by epidemiologic methods is necessary. A latency period is usual before symptoms become prominent (up to several years for textile workers, months for typical allergens).

Physical examination, chest x-ray, and routine spirometric testing can establish the hallmarks of obstructive lung disease (wheezing, hyperinflation, reversible airways obstruction) but have no specificity for work-relatedness. They can and should be used to establish degree of impairment. Peak flow measurements, accompanied by diary records, can establish diurnal variations, work-relatedness, and response to a changing environment. Airway lability with consistent variation in peak flow of more than 20% between AM and PM measurements suggests the kind of diurnal variations that occur with asthma. If time is spent explaining and demonstrating the maneuver to the patient, consistent and reliable readings can be obtained in spite of the effort-dependence of this measurement.

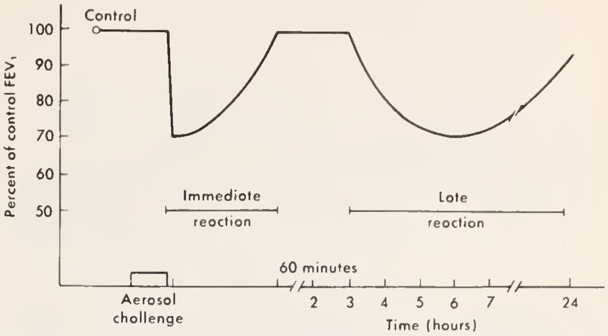


Fig. 1. Airway responses to inhaled antigen. Dual response, involving immediate and late reactions. *Adapted from* Lehnert B, Schachter EN. Respiratory pharmacology. St. Louis: CV Mosby, 1980, Fig. 5–12).

Nonspecific airways challenge (40) testing is frequently used to establish the presence of airway irritability in workers. Abnormal results help to confirm the presence of asthma. The development of severe persistent airway hyperreactivity in the setting of an acute exposure has been used to define RADS. Nevertheless, these tests have serious shortcomings as validation of work-relatedness of disease. Negative responses do not exclude occupational causation (41, 42). Hyperreactivity may wane in some workers removed from an offending exposure (43), but not in all, particularly when exposure is great (RADS) or sustained (44).

Skin testing and serum precipitins may be useful in characterizing occupational airway disease. Atopy can be established by demonstrating positive skin reactions to 2 or more common allergens. Skin testing with specific occupational antigens may corroborate a highly suspicious case but should not be considered definitive proof. If the antigen—for example, platinum salts (45)—is likely to be encountered only in the workplace, the specificity of the test increases. Skin tests with antigens from animal products (46), castor bean extract (47), coffee (48), and flour (49) also produce skin reactions in sensitized workers but are not as specific because of the ubiquity of these

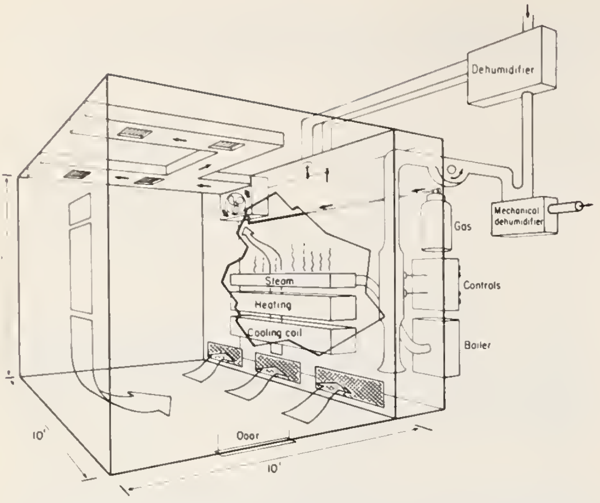


Fig. 2. Environmental chamber facility (27 $m^3$). Design allows investigator to control temperature, humidity, and ventilation. Internal air circulation permits rapid and uniform distribution of any tested agent. Physiologic testing is performed within chamber. *Adapted from* Witek TJ Jr, Schachter EN, eds. Advances in respiratory pharmacology. (ref 40), Fig. 9.

agents. The demonstration of IgE antibodies has also been used to confirm sensitization, particularly in response to antigens of high molecular weight such as laboratory animals (50), flour (49), coffee (51). IgE antibodies to compounds of low molecular weight—isocyanates (52), wood dust (53), nickel (54)—have not consistently been found in symptomatic workers.

Specific inhalation testing is frequently used to confirm that an airway response is linked to a single workplace agent rather than to a combination of agents. Unlike evaluation in the natural setting of the workplace, laboratory challenges are performed in controlled environments such as an environmental chamber (Fig. 2). Challenges may be performed with irritants (gases) or allergens (antibiotics, extracts of animals or plant products). Because delayed responses may occur with such challenge, it has been recommended that they be performed in specialized facilities and that the patient be monitored for at least 24 hours after challenge. The safety of testing personnel must also be considered.

Specific challenge is indicated when

- the clinical and laboratory features of the patient's airway disease do not permit easy or sufficiently definitive association with the workplace agent
- an agent not previously characterized is suspected
- the effects of several agents in the worker's environment need to be distinguished

Specific bronchial provocation attempts to duplicate the exposure and the dose experienced by the worker in the workplace. A number of different approaches have been described (55). Provocation with traditional antigen extracts prepared for nebulization via a standard aerosol generator using the protocol of Chai and co-workers (56) has frequently been criticized as delivering an unrealistic challenge not reflective of the workplace. Occupational materials such as grain dust or wood dust can be inhaled as a challenge by pouring workplace dust from one pan to another in an otherwise clean chamber. Fine powders such as antibiotics or platinum salts are mixed with a vehicle such as dried lactulose. Soldering can be performed in such an environment to recreate workplace vapors associated with colophony.

The dose of standard allergens producing immediate hypersensitivity can be gauged by skin testing. A subthreshold dose as evaluated by skin testing can be administered as an aerosol and the dose varied either by concentration or by length of exposure. If skin testing does not yield immediate reactions, the aerosol challenge dose must be evaluated empirically by estimated work exposures or knowledge of the agent's ability to cause bronchospasm.

Responses have been delayed not only to the evening or the night following provocation but also to subsequent days (57, 58). Thus follow-up of patients using peak flow meters is warranted beyond 24 to 48 hours.

Bronchial alveolar lavage has been used to characterize the inflammatory response induced by occupational agents (59) but is not at this point a diagnostic maneuver of established value.

Workplace studies, exclusive of monitoring lung function in the workplace, for a given worker are most appropriately performed in the context of an epidemiological survey. Limited surveys may be helpful in gaining a better understanding of the relationship between the worker and the suspected agent and better characterizing the processes by which a worker is exposed.

**Therapy.** The ideal therapeutic strategy consists of removing the person from the offending agent, since the disease may involve sensitization, and exposure to even low levels of the agent may aggravate and perpetuate the disease. Guides or constraints to management options include:

- industry size
- prevalence of the occupational disease
- federal regulations: dust level standards (OSHA), air pollution standards (EPA)

- mandated medical surveillance, e.g., annual screening with questionnaire and spirometry
- job seniority, security, and transfer policies and possibilities
- medical screening of new workers
- control of smoking
- improving ventilation
- medical therapy
- respirators and protective devices

Pharmacologic therapy to relieve symptoms of occupational airway disease is similar to that for nonoccupational asthma, although some drugs such as disodium cromoglycate may have a special role in workplace-related syndromes.

### Two (of Two Hundred) Syndromes

Although it is not possible to discuss in detail here the 200 or more known agents responsible for occupational airway disease, two syndromes with particularly striking characteristics are discussed.

**Byssinosis.** Byssinosis is an occupational lung disease associated with the inhalation of cotton and other textile dusts (60, 61). Over 800,000 textile workers in the United States are at risk of developing byssinosis; cotton processing is by far the largest industry whose workers are at risk. Epidemiologic studies have now extensively documented both acute and chronic effects on the airways of textile workers.

In the early phases of byssinosis, acute reversible symptoms such as wheezing, chest tightness, and shortness of breath accompany reversible changes in lung function (62). "Monday dyspnea" is the term used to describe these symptoms, which occur on the first day back at work after an absence. The disease may progress to a stage in which symptoms are present throughout the work week and may eventually result in severe respiratory disability (63), a stage estimated to occur in 7% of older cotton textile workers (64). The characterization of the disease has primarily been based on respiratory symptoms as well as the clinical grading of Schilling (65):

Grade 0. No particular symptoms on first day of working week.
Grade 1/2. Occasional chest tightness or irritation of respiratory tract (e.g., cough) on first day of working week.
Grade 1. Chest tightness on every first day of working week.
Grade 2. Chest tightness on first and other days of working week.
Grade 3. Grade 2 symptoms accompanied by evidence of permanent incapacity.

Lung function measurements are used clinically and epidemiologically to evaluate the severity of illness and to document acute changes over a work shift. These measurements form the basis for medical surveillance currently mandated in the cotton textile industry.

Inherent in Schilling's clinical grading of byssinosis is the distinction between acute reversible bronchospasm and chronic, debilitating airflow obstruction. Although the acute response has been characterized by across-shift changes in lung function and the chronic response by the presence of chronic bronchitis, impaired lung function, or both, the transition between the two stages of the disease is still poorly understood. Prospective studies in cotton textile workers have documented an accelerated decline in lung function. Molyneux, Tombleson, and Berry (66–68) conducted some of the earliest prospective studies on byssinosis using a cohort of Lancashire workers in 14 cotton mills and two artificial-fiber mills (controls). A total of 1300 cotton workers and 227 fiber workers were examined in the original cross-sectional study. The prospective analysis studied 595 individuals. Follow-up examinations were conducted at 6-month intervals over the 3-year period from 1963 to 1966. No observable dose-response relationship between dust levels and chronic bronchitis were noted. Nevertheless, the average annual decline in forced expiratory volume in 1 second ($FEV_1$) was 54 mL/year for cotton workers and 32 mL/year for workers in the artificial-fiber mills. In a study from Bombay reported by Kamat and co-workers (69), the level of symptom prevalence was associated with length of service in the cotton mills. In addition, mean annual decreases in $FEV_1$ were greater in general with increasing years worked in both symptomatic and asymptomatic individuals. Similarly, a trend in yearly $FEV_1$ change was noted by dust exposure levels.

In 1973, Fox and co-workers (70) reported results from workers in 35 Lancashire area mills. Approximately 2300 workers were examined from 1966 to 1968, and 2500 workers from 1968 to 1970. Of these workers, 886 were common to both studies. Those workers who progressed from low grades of byssinosis to higher grades over the period of follow-up had greater annual declines than those who began at higher grades and improved to lower grades at reexamination. For example, male workers progressing from Grade 1 to Grade 2 byssinosis had a mean annual decline of $FEV_1$ of 179 mL, as opposed to a decline of 62 mL for those remaining at Grade 1. However, the 3-year follow-up study by Berry and colleagues failed to confirm this pattern, as workers progressing from grades "Nil or 1/2" to grades "1 or 2" lost 37 mL/year on average, whereas those workers remaining at "Nil or 1/2" lost 56 mL/year.

In one of the few prospective studies to be performed in the United States, Beck and associates (63) studied active and retired textile workers 45 years of age and older in Columbia, South Carolina. Community controls were taken from two towns in Connecticut and a geographically neighboring town, Winnsboro, South Carolina. A total of 383 cotton textile workers and 277 controls were studied longitudinally over a 6-year period. Chronic bronchitis developed in 16% of the cotton textile workers compared with 1% of controls. Male textile workers lost 42/mL per year of $FEV_1$ as compared with 25 mL/year for male controls. Retired cotton textile workers had more symptoms and impairment (lower percentage of predicted $FEV_1$) than did active workers.

In a separate report Beck et al. (71) noted that both smoking and cotton dust influenced lung function, and that these effects were additive (for example, in male workers 45 years or older, an average of 0.407 L loss was attributed to cotton dust, with a corresponding smoking effect of 0.368 L). Possible dose-response relationships from these prospective data could not be fully evaluated, as dust measurements were not available. Similar findings have recently been reported in a 10-year follow-up of workers in a very dusty mill in Yugoslavia (72).

The essential dose-relatedness of byssinosis was established in the 1970s by Merchant et al. (73), who showed an increasing prevalence of byssinosis grades with higher dust levels. These relationships form the basis for the current stringent standards established for the regulation of the cotton textile industry in the United States. Additional refinement of these standards may come as more specific surrogates of the byssinotic effect are characterized—for example, endotoxin (74)—and our understanding of the airway active agent(s) in cotton dust develops.

**Soy-Dust Asthma.** Successful control of point-source epidemics are the result of both serendipity and painstaking research. Characteristically, they proceed from astute observation—in general by primary care physicians—of unusual clusterings of disease, to the careful detective work necessary to define patterns of the disease in the affected population, to the discovery of a specific cause of the outbreak, and ultimately to control measures. Such was the history of cholera epidemics in London in the nineteenth century, with the dramatic story of the Broad Street pump; such is the story of the asthma epidemic in Barcelona in the 1980s.

Beginning in 1981, "outbreaks" of asthma were identified in Barcelona, Spain, by a sudden and remarkable increase in asthma emergency room visits on otherwise unremarkable days. In 1983, Ussetti and co-workers (75) reported 6 outbreak days with characteristic epidemic features. Asthmatics came to emergency rooms within a narrow period, reporting paroxysmal symptoms. A subsequent report noted that high ambient levels of nitrogen dioxide were associated with at least one epidemic (76). In 1984, monitoring of emergency room admissions in four city hospitals, representing the majority of Barcelona's emergency care facilities, was established. Analysis of subsequent outbreaks confirmed that on certain days adult emergency room visits for asthma reached levels far in excess of that expected (77, 78). Study of the geographic distribution of cases showed a clustering of the affected individuals in the proximity of Barcelona's harbor area. Meteorologic, aero-allergen, and pollutant monitoring failed to identify any known causal agent. However, because the highest risk areas were near the harbor and industrial areas, a point-source epidemic was considered likely as the cause (77).

Analysis of subsequent episodes in 1985 and 1986 confirmed the geographic predilection of the outbreaks (78). In 1987 it was noted that at least three epidemics occurred on days on which soybeans were unloaded in the harbor. Retrospective review of previous outbreaks revealed a high degree of geographic correlation on 13 epidemic days; in particular, the association was striking for the unloading of soybean in Silo A, one of two major silos in the port area. Silo A was noted not to have an adequate filtration system to protect the ambient air (79).

In a separate study involving serum from asthmatics affected during the prior episodes and from matched controls, 74% of the affected individuals showed high IgE titers specific for soybean, whereas only 5% of the controls showed high soybean-specific IgE titer findings (80). These serologic studies confirmed the findings from the epidemiologic studies, which suggested an immunologic mechanism for outbreaks. Filters were subsequently fitted to silo A, and epidemic asthma in Barcelona has since been controlled.

## Conclusion and Summary

Occupational airway diseases form a significant proportion of airway disease, but the occupational associations frequently go unnoticed. An understanding of the well-delineated airway syndromes occurring as a result of occupational exposures and a routine inquiry as to possible workplace or environmental exposures will help the primary care physician and the specialist identify potentially preventable or reversible disease.

### References

1. Ramazzini B. Diseases of workers (1713). Tr. from Latin by Wilmer Cave Wright. New York: Hafner Publishing Co. 1964, 13.

1a. Bouhuys A, Beck GJ, Schoenberg JB. Epidemiology of environmental lung disease. Yale J Biol Med 1979; 52:191–210.

2. Beck GJ, Maunder LR, Schachter EN. Cotton dust and smoking effects on lung function in cotton textile workers. Am J Epidemiol 1984; 119:33–43.

3. Brooks SM, Weiss MA, Bernstein IL. Reactive airways dysfunction syndrome (RADS). Chest 1985; 88:376–383.

4. Jones GR, Proudfoot AT, Hall JI. Pulmonary effects of acute exposure to nitrous fumes. Thorax 1973; 28:61–65.

5. Brooks SM, Kalica AR. NHLBI Workshop Summary: strategies for elucidating the relationships between occupational exposures and chronic air-flow obstruction. Am Rev Respir Dis 1987; 135:268–273.

6. Fish JE. Occupational asthma: a spectrum of acute respiratory disorders. JOM 1982; 24:379–386.

7. Weiss KB, Wagner DK. Asthma surveillance in the United States: a review of current trends and knowledge gaps. Chest (suppl) 1990; 98:179s–184s.

8. Weiss ST. A framework for assessing impairment from asthma. Chest (suppl) 1990; 98:225s–231s.

9. Schachter EN, Doyle CA, Beck GJ. A prospective study of asthma in a rural community. Chest 1984; 85:623–630.

10. Beck GJ, Doyle CA, Schachter EN. A longitudinal study of respiratory health in a rural community. Am Rev Resp Dis 1982; 125:375–381.

11. Zuskin E, Skuric Z, Kanceljak B, Prokrajac D, Schachter EN, Witek TJ. Respiratory symptoms and ventilatory capacity in soy workers. Am J Indust Med 1988; 14:157–165.

12. Zuskin E, Skuric Z, Kanceljak B, Prokrajac D, Schachter EN, Witek TJ. Respiratory findings in spice factory workers. Arch Env Health 1988; 43:335–339.

13. Zuskin E, Skuric Z, Kanceljak B, Prokrajac D, Schachter EN, Witek TJ. Respiratory symptoms and lung function in furriers. Am J Indust Med 1988; 14:189–196.

14. Zuskin E, Mataija M, Prokrajac D, Schachter EN, Witek TJ. Respiratory function in animal food processing workers. Am J Indust Med 1989; 16:179–187.

15. Zuskin E, Kanceljak B, Prokrajac D, Schachter EN, Witek TJ Jr. Respiratory symptoms and lung function in hemp workers. Br J Indust Med 1990; 47:627–632.

16. Speizer FE. Asthma and persistant wheeze in the Harvard six cities study. Chest 98 (suppl): 1990; 191s–195s.

17. Korn RJ, Dockery DW, Speizer FE, Ware JH, Ferris BG Jr. Occupational exposures and chronic respiratory symptoms. Am Rev Respir Dis 1987; 136:298–304.

18. Schwartz DA, Wakefield DS, Fieselmann JF, Berger-Wesley M, Zeitler R. The occupational history in the primary care setting. Am J Med 1991; 90:315–319.

19. Brooks SM. Occupational asthma. Chest 87 (suppl) 1985; 218s–225s.

20. Blanc P. Occupational asthma in a national disability survey. Chest 1987; 92:613–617.

21. Evan R, Mullaly DI, Wilson RW, et al. National trends in morbidity and mortality of asthma in the U.S. Chest 1987; 91 (suppl): 65s–73s.

22. Chan-Yeung M, Ashley MJ, Corey P, Willson G, Dorken E, Grzybowski S. A respiratory survey of cedar mill workers: I. prevalence of symptoms and lung function abnormalities. J Occup Med 1978; 20:328–332.

23. Mitchell ML, Gandevia B. Respiratory symptoms and skin sensitivity in workers exposed to proteolytic enzyme in detergent industry. Am Rev Respir Dis 1971; 104:1–12.

24. Schachter EN, Witek TJ, Beck GJ, Colice G, Hosein HR, Leaderer BP, Cain W. Dose response effect of low concentrations of sulfur dioxide on airway function. Arch Envir Health 1984; 39:34–42.

25. Schachter EN, Witek TJ, Brady DJ, Tosun T, Beck GJ, Leaderer BP. A study of respiratory effects from exposure to 2.0 ppm formaldehyde in occupationally exposed workers. Env Res 1987; 44:188–205.

26. Hendrick DJ, Love DJ. Formalin asthma in hospital staff. Br M J 1975; 1:607–608.

27. Pepys J, Pickering CAC, Hughes EC. Asthma due to inhaled chemical agents: complex salts of platinum. Clin Allerg 1972; 2:391–396.

28. Fawcett IW, Newman-Taylor AJ, Pepys J. Asthma due to inhaled chemical agents: epoxy resin systems containing phthalic acid anhydride, trimellitic acid anhydride and triethylene tetramine. Clin Allerg 1977; 7:1–14.

29. Zeiss CR, Patterson R, Pruzansky JJ, Miller M, Rosenberg M, Levitz D. Trimellitic anhydride-induced airways syndromes: clinical and immunologic studies. J Allerg Clin Immunol 1977; 60:96–103.

30. Pepys J, Pickering CAC, Breslin ABX, Terry DJ. Asthma due to inhaled chemical agents: toluene diisocyanate. Clin Allerg 1972; 2:225–236.

31. Butcher BT, Salvaggio JE, Weill H, Ziskind MM. Toluene diisocyanate (TDI) pulmonary disease: immunologic and inhalation challenge studies. J Allergy Clin Immunol 1970; 58:89–100.

32. Chan-Yeung M. Immunologic and nonimmunologic mechanisms in asthma due to western red cedar (*Thuja plicata*). J Allerg Clin Immunol 1982; 70:32–37.

33. Davies RJ, Hendrick DJ, Pepys J. Asthma due to inhaled chemical agents: ampicillin, benzylpenicillin, 60-aminopenicillanic acid and related substances. Clin Allerg 1974; 4:227–247.

34. Coutts II, Dally MB, Newman-Taylor AJ, Pickering CAC, Horsfield N. Asthma in workers manufacturing cephalosporins. Br Med J 1981; 283:950.

35. Diaz P, Gonzales MC, Gelleguillos FR, et al. Leukocytes and mediators in bronchoalveolar lavage during allergen-induced late phase asthmatic reactions. Am Rev Respir Dis 1989; 139:1383–1389.

36. Lam S, Le Riche J, Phillips D, Chan Yeung M. Cellular and protein changes in bronchial lavage fluid after late asthmatic reaction in patients with red cedar asthma. J Allerg Clin Immunol 1987; 80:44–50.

37. Fabbri LM, Boschetto P, Zocca E, et al. Bronchoalveolar neutrophilia during late asthmatic reactions induced by toluene diisocyanate. Am Rev Resp Dis 1987; 136:36–42.
38. Butcher BT, Salvaggio JE, Weill H, Zuskind MM. Toluene diisocyanate (TDI) pulmonary disease: immunologic and inhalation challenge studies. J Allerg Clin Immunol 1970; 58:89–100.
39. Gallagher JS, Tse CST, Brooks SM, Bernstein IL. Diverse profiles of immunoreactivity in toluene diisocyanates (TDI) asthma. JOM 1981; 23:610–616.
40. Witek TJ, Schachter EN. Bronchial provocation to assess airway responsiveness. In: Witek TJ, Schachter EN, eds. Advances in respiratory care pharmacology. Philadelphia: J. P. Lippincott, 1989, 88–103.
41. Hargreave FE, Ramsdale EH, Pugsley SO. Occupational asthma without hyperresponsiveness. Am Rev Resp Dis 1984; 30:513–515.
42. Burge PS. Nonspecific bronchial hyperreactivity in workers exposed to toluene diisocyanate, diphenyl methane diisocyanate and colophony. Eur J Respir Dis 1982; 123:91–96.
43. Cartier A, Pineau L, Malo JL. Monitoring of maximum expiratory peak flow rates and histamine inhalation tests in the investigation of occupational asthma. Clin Allerg 1984; 14:193–196.
44. Chan-Yeung M, MacLean L, Paggiero PL. A follow-up of 232 patients with occupational asthma due to western red cedar (*Thuja plicata*). J Allerg Clin Immunol 1987; 80:279–284.
45. Pepys J, Pickering CAC, Hughes EG. Asthma due to inhaled chemical agents: complex salts of platinum. Clin Allerg 1972; 2:391–396.
46. Cockroft A, Edwards J, McCarthy P, Anderson N. Allergy in laboratory animal workers. Lancet 1981; 1:827–830.
47. Davison AG, Britton MG, Forrester JA, Davies RJ, Hughes DTD. Asthma in merchant seamen and laboratory workers caused by allergy to castor beans: analysis of allergens. Clin Allerg 1983; 13:553–561.
48. Zuskin E, Valic F, Kanceljak B. Immunological and respiratory changes in cotton workers. Thorax 1981; 36:9–13.
49. Thiel H, Ulmer WNT. Baker's asthma development and possibility of treatment. Chest 78 (suppl): 1980; 400–405.
50. Newman Taylor AJ, Longbottom JL, Pepys J. Respiratory allergy to urine proteins of rats and mice. Lancet 1977; 2:847–849.
51. Karr RM. Bronchoprovocation studies in coffee worker's asthma. J Allerg Clin Immunol 1979; 64:650–654.
52. Butcher BT, Salvaggio JE, Weill H, Ziskind MM. Toluene diisocyanate (TDI) pulmonary disease: immunologic and inhalation challenge studies. J Allerg Clin Immunol 1970; 58:89–100.
53. Bush RK, Clayton D. Asthma due to Central America walnut (*Juglans olanchana*) dust. Clin Allerg 1983; 13:389–394.
54. Block G, Chan Yeung M. Asthma induced by nickel. JAMA 1982; 247:1600–1602.
55. Newman Taylor AJ, Davies RJ. Inhalation challenge testing. In: Weill H, Turner-Warwick M. Occupational lung diseases. New York: Marcel Dekker, 1981, 143–157.
56. Chai H, Farr RS, Frochlick LA, et al. Standardization of bronchial inhalation procedures. J Allerg Clin Immunol 1975; 56:323–327.
57. Newman Taylor AJ, Davies RJ, Hendrick DJ, Pepys J. Recurrent nocturnal asthmatic reactions to bronchial provocation tests. Clin Allerg 1979; 9:213–219.
58. Davies RJ, Green M, Schofield NM. Recurrent nocturnal asthma after exposure to grain dust. Am Rev Resp Dis 1976; 114:1011–1019.
59. Merrill WW, Buck M, Cooper JAD, Schachter EN. Relationship of the ventilatory response to cotton bract extract and the cells and proteins of the lung. Chest 1987; 91:44–48.
60. Bouhuys A, Zuskin E. Chronic respiratory disease in hemp workers: a follow-up study, 1967–1974. Ann Intern Med 1976; 84:398–405.
61. Bouhuys A, Schoenberg JB, Beck GJ, Schilling RSF. Epidemiology of chronic lung disease in a cotton mill community. Lung 1977; 154:167–187.
62. Zuskin E, Valic F, Bouhuys A. Byssinosis and airway responses due to exposure to textile dust. Lung 1976; 154:17–24.
63. Beck GJ, Schachter EN, Maunder LR, Schilling RSF. A prospective study of chronic lung disease in cotton textile workers. Ann Intern Med 1982; 97:645–651.
64. Bouhuys A, Beck GJ, Schoenberg JB. Priorities in the prevention of chronic lung disease. Lung 1979; 156:129–148.
65. Schilling RSF. Worldwide problems of byssinosis. Chest 1981; 79:3S–5S.
66. Molyneux MKB, Tombleson JBL. An epidemiological study of respiratory symptoms in Lancashire mills, 1963–66. Br J Indust Med 1970; 27:225–234.
67. Berry G, McKerrow CB, Molyneux MKB, et al. A study of the acute and chronic changes in ventilatory capacity of workers in Lancashire cotton mills. Br J Indust Med 1973; 30:25–36.
68. Berry G, Molyneux MKB, Tombleson JBL. Relationship between dust level and byssinosis and bronchitis in Lancashire cotton mills. Br J Ind Med 1974; 31:18–27.
69. Kamat SR, Kamat GR, Salpekar VY, Lobo E. Distinguishing byssinosis from chronic obstructive pulmonary disease: results of a prospective 5 year study of cotton mill workers in India. Am Rev Respir Dis 1981; 124:31–40.
70. Fox AJ, Tombleson JBL, Watt A, Wilkie AG. A survey of respiratory disease in cotton operatives: Part I. symptoms and ventilation test results. Br J Indust Med 1973; 30:42–47.
71. Beck GJ, Maunder LR, Schachter EN. Cotton dust and smoking effects on lung function in cotton textile workers. Am J Epidemiol 1984; 119:33–43.
72. Zuskin E, Ivankovic D, Schachter EN, Witek TJ. A ten year follow-up study of cotton textile workers. Am Rev Resp Dis 1991; 143:301–305.
73. Merchant JA, Lumsden JC, Kilburn KH, O'Fallon WM, Ujda JR, Germino VH, Hamilton JD. Dose response studies in cotton textile workers. J Occup Med 1973; 15:222–238.
74. Castellan RM, Olenchock SA, Kinsley KB, Hankinson JL. Inhaled endotoxin and decreased spirometric values. N Engl J Med 1987; 317:605–610.
75. Ussetti P, Roca J, Augusti AGN, Montserrat JM, Rodriguez-Roisin R, Augusti-Vidal A. Asthma outbreaks in Barcelona. Lancet 1983; 2:280–281.

76. Ussetti P, Roca J, Augusti AGN, Montserrat JM, Rodriguez-Roisin R, Augusti-Vidal A. Another asthma outbreak in Barcelona. Lancet 1984; 1:156.
77. Anto JM, Sunyer J. Barcelona Asthma Collaborative Group: a point source asthma outbreak. Lancet 1986; 1:900–903.
78. Hendrick DJ. Asthma: epidemics and epidemiology. Thorax 1989; 44:609–613.
79. Anto JM, Sunyer J, Rodriguez-Roisin R, Suarez-Cervera M, Vasquez L. Community outbreaks of asthma associated with inhalation of soybean dust. N Engl J Med 1989; 320:1097–1102.
80. Sunyer J, Anto JM, Rodrigo MJ, Morell F. Case-control study of serum immunoglobulin-E antibodies reactive with soybean in epidemic asthma. Lancet 1989; 1:179–182.

# The Role of Atopy in Asthma

KIRK SPERBER, M.D.

THE DEVELOPMENT of airway obstruction is responsible for the clinical manifestations of asthma (1). This obstruction is believed to be initiated and propagated by inflammatory events (2, 3) that involve the recruitment of proinflammatory cells like eosinophils and neutrophils (4) that disrupt normal airway function. The mechanisms leading to this inflammation in asthma are multiple and complex. Alteration of normal lung function leads to bronchospasm and obstruction (2) as well as mucus hypersecretion and plugging. The presence of these inflammatory cells and their mediators produces alterations in bronchial epithelial cells (5) that lead to cellular destruction and increased bronchial smooth muscle hyperreactivity.

**Inflammation.** Evidence of the prominent role of inflammation in asthma comes from morphologic studies of asthmatic patients demonstrating that bronchial inflammation is present in all stages of disease, although it is most prominent in severe asthma (4, 6). Inflammation causes a marked hyperresponsiveness of the airway (defined as an exaggerated bronchoconstrictive response) to a wide variety of physical, chemical, or pharmacologic agents like methacholine or histamine, so that the threshold for bronchospasm is reduced (7). Airway hyperresponsiveness is a constant feature of asthma; the level of hyperresponsiveness (8) is related to the degree of underlying inflammation and usually correlates with the severity of disease as well as medication requirements (9).

## Allergy

Although many triggers can initiate airway inflammation, it has become increasingly apparent that aeroallergens contribute significantly. Allergic IgE-mediated reactions in the airway are important for two reasons in asthma: immediate reactions to aeroallergens (0–60 min) can elicit bronchospasm, and immediate reactions provoke late phase reactions, which occur 4 to 6 hours after the initial response (10). These reactions are caused when IgE is cross-linked on the surface of mast cells through interaction with an allergen. This results in immediate histamine release and a late phase reaction 4–6 hours later (11).

The activated mast cell releases a variety of chemoattractants, including neutrophil, eosinophil, and mononuclear cell chemotactic factors, which cause the accumulation of these cells at the site of the activation (12, 13). The late phase reactions provoke increased bronchial hyperresponsiveness, which can persist for weeks after the initial exposure to allergen. The presence of the neutrophils, eosinophils, and mononuclear cells has been documented in asthmatic patients in several studies (12, 13). In bronchial-alveolar-lavage fluid from asthmatics who were challenged with antigen in the airway, eosinophils and mast cells are increased at 6 hours after challenge (12, 13). Increases in eosinophils and mast cells have also been observed in sputum during asthma exacerbations, in contrast to levels in normals, acute bronchitics, and smokers, whose sputum is characterized by an increase in mononuclear cells, not eosinophils or mast cells (14).

Although allergens are important in asthma, they must enter into the lower airway in order to provoke the inflammatory response. Since most pollen particles 2 to 60 μm and are removed by the efficient anatomic barriers in the upper airway, entry into the lower airway must occur by alternate means. Several studies have demon-

---

From the Division of Clinical Immunology, Department of Medicine, The Mount Sinai Medical Center, One Gustave Levy Place, New York, NY. Address correspondence and reprint requests to Kirk Sperber, M.D., Division of Clinical Immunology, Box 1089, One Gustave Levy Place, New York, NY 10029.

strated that fragments of the pollen particle containing active allergen, not the entire pollen grain, associate with airborne particles like rain drops, which can gain access to the lower airways (15–17) to interact with IgE bound to mast cells.

Clinical evidence that IgE-mediated allergic reactions are important in asthma comes from several studies that have either documented positive skin tests, indicating the presence of an IgE antibody to a particular allergen, for aeroallergens, or shown elevated IgE levels with asthma. It is not possible to predict who will be asthmatic or who will have allergic rhinitis by skin test. In a recent large 10-year multicentered national study (18), asthma associated with allergic rhinitis was found to be highly related to allergic skin reactivity for all age groups among nonsmokers. Additionally, there was an association of asthma without allergic rhinitis and positive skin tests, although this was less statistically significant. There were technical difficulties with these studies in that the skin testing used a panel of aeroallergens that did not include those associated with asthma, including house dust, dust mite, cat, and cockroach.

Several other studies link IgE-mediated reactions (atopy) and asthma, including one by Burrows et al. (19), that found elevated IgE levels in a large group of patients coming to emergency rooms for asthma in Tucson, Arizona, and a study by Kalliel et al. (20), who reported that 60% of adult asthmatics attending a pulmonary clinic were atopic. Pollart et al. (21) have demonstrated positive skin tests for house dust mites in adult asthmatics coming to emergency rooms for treatment; Barbee et al. (22) and Licorish et al. (23) have reported similar findings of positive skin tests for mold spores and asthma. We at the Mount Sinai Medical Center in New York City, N.Y., have similar data from predominantly black and Hispanic patients in Spanish Harlem who are attending the asthma clinic at our institution; 80% of all the clinic patients had at least three positive skin tests for aeroallergens. The most common allergens in our group included cockroach, mite, cat, and dog (Table).

TABLE

*Skin Test Profile, 93 Patients Allergic to Indoor Panel of Allergens (%)*

| Allergen | Prick | I.D. | Neg. |
|---|---|---|---|
| Dust | 29 | 69 | 2 |
| Mite | 33 | 57 | 10 |
| Dog | 30 | 51 | 19 |
| Cat | 34 | 42 | 24 |
| Cockroach | 20 | 45 | 35 |
| Ragweed | 39 | 25 | 36 |
| Fall weed | 31 | 30 | 39 |
| Grass | 36 | 19 | 45 |
| Plaintain | 19 | 35 | 46 |
| Tree | 19 | 33 | 48 |
| Mold | 10 | 36 | 54 |
| Mean | 3.0 | 4.4 | 3.6 |

Mean number of positive skin tests per patient is 7.4.

**Age Variation.** The importance of allergy in asthma varies in different age groups. In infants, viral infections are the main asthma triggers; allergens play less of a role because of the time required for sensitivities to develop (24). Food allergies generally do not play an important role in asthma in infants. In children there is considerable evidence that allergy can influence the persistence and severity of the disease. Several reports correlate the severity of childhood asthma with the number of positive skin tests for aeroallergens (25, 26). Additionally, children with multiple positive allergy skin tests are more likely to have daily rather than intermittent asthma because of the presence of chronic allergic inflammation (27).

In adults atopy is important in asthma, especially in those patients whose disease began before 30 years of age, although patients over 30 can also experience atopic reactions for the first time. Exposure to occupational aeroallergens like phthalic anhydride (28) from plastic, phenylglycine from auto paint (29), and trimellitic anhydride (30) from epoxy resins can also influence the development of bronchial hyperreactivity and asthma.

In contrast to children, in adults the number of positive skin tests does not appear to correlate to the severity of asthma (31). Food allergy in adults, as in children, does not appear to be a common asthma trigger.

**Aspirin Provocation.** Aspirin sensitivity can provoke asthma on a nonimmune basis. This sensitivity is relatively common in asthmatics, although wide variation has been reported in the incidence in various asthmatic populations (32). Aspirin sensitivity is associated with nasal polyposis and chronic sinusitis. In contrast to IgE-mediated inflammation, details regarding pathogenesis are not clearly elucidated.

Many theories attempt to explain this disorder. The most widely accepted is blockade of the cyclooxygenase enzyme by aspirin which leads to diversion of arachidonic acid products into the lipoxygenase pathway leading to the generation of powerful bronchoconstrictors like $LTC_4$, $LTD_4$ and $LTE_4$ (33).

Avoidance of aspirin and other nonsteroid antiinflammatory drugs, surgical resection of the nasal polyps, and early treatment of the sinusitis are essential to manage this disorder.

### Diagnosis of Allergy

The diagnosis of allergy in an asthmatic population can sometimes be difficult, especially when allergen exposure occurs continuously (house dust, dust mite, and animal dander) rather than seasonally. The symptoms caused by this indoor panel of allergens may be aggravated by housecleaning and generally are worse in the winter months (34). Levels of indoor allergens in the United States can vary considerably owing to the great diversity in winter temperatures and levels of humidity. Although these factors were thought only to influence levels of seasonal pollens, they also affect levels of house-dust mites and cockroaches, as well as indoor and outdoor molds (34).

Seasonal pollens, including ragweed, trees, and grass, are also associated with seasonal asthma (35, 36). Whether asthma attacks occur depends on several factors, including the concentration of pollen to which an individual is exposed and the degree of exposure (37, 38). Thus, high concentrations of airborne pollens like ragweed, trees, and grass can become a major cause of lower airway inflammation, which can lead to bronchospasm.

Pollen can be collected by several methods (39); using air sampling and comparative skin testing, a correlation between pollen count, degree of skin reactivity, and clinical symptoms can be demonstrated (36, 40). Different pollens predominate at different times of the year. Tree pollens are present from February through May, grasses from May to June, weeds from August to October, and outdoor molds throughout the warm months (41).

**Demonstrating IgE Antibodies to Aeroallergens.** The presence of IgE antibodies directed against aeroallergens can be demonstrated by 3 methods: skin testing, radioallergosorbent (RAST) testing, and the leukocyte histamine release test (42).

Skin testing allows for the interaction of IgE bound to mast cells with allergen introduced into the skin either by pricking the skin or by intradermal injection. This method is valid because the same allergen-IgE interaction that occurs in the upper and lower airways can be reproduced in the skin. A wheal-and-flare response typical of histamine release from mast cells indicates a positive reaction. Each study must include a positive histamine control and a negative normal saline control. In several chronic inflammatory skin conditions, including dermatographism and atopic eczema, skin testing is difficult to interpret. Furthermore, antihistamine therapy causes skin tests to be falsely negative by blocking histamine receptors.

RAST testing measures specific IgE antibodies against allergens (42). It is a two sided immunometric test where a specimen to be tested is reacted in the first stage with an allergen that is coupled to a solid phase material like sephadex beads, filter paper discs, agarose beads or microtiter wells. In the second stage, monoclonal anti-IgE antibodies which are either radiolabeled or enzyme linked are used to measure the amount of bound IgE immunoglobulin. Standard concentrations of specific anti-allergen IgE antibody are included in each assay to generate a standard curve to measure the quantity of IgE antibody in each sample.

Histamine release from peripheral blood basophils can also be used to detect allergen-specific IgE antibodies (42). In these studies, peripheral blood leukocytes are isolated by dextran sedimentation and resuspended in buffer containing $Ca^{2+}$ and $Mg^{2+}$. Differing concentrations of allergy extract are added to a fixed number of cells for varying periods of time. Histamine is then assayed from the culture supernatant.

The results of these three tests can identify relevant allergens. Although either a positive skin test or RAST or histamine release assay is necessary to diagnose allergy, one may encounter positive skin tests without clinical relevance (43). The prevalence of false-positive skin tests makes the patient's clinical history extremely important; skin tests should be used only to confirm the presence of allergy. However, in at least one study positive skin tests in the absence of clinical disease predicted the development of allergic diathesis in college students followed for 10 years (44, 45). The incidence of false-positive RASTs is extremely low; it is estimated that up to 15% of nonatopic patients can have histamine release in response to allergens. In vitro laboratory testing can also serve as a useful replacement for skin testing. However, the increased cost and decreased sensitivity still makes skin testing the preferred diagnostic modality (42).

### Treatment

Although asthma is a complex disease in which allergic factors are not the only ones responsible for bronchial inflammation, they do play a major role (1, 2, 11, 12). It has been demonstrated (46) that during pollen season grass-sensitive asthmatics have increased bronchial hyperresponsiveness. McAllen et al. have shown that the threshold dose to provoke nonspecific hyperreactivity in response to inhaled methacholine was significantly increased after immunotherapy. Specific therapy, then, aimed at reducing allergen-specific IgE antibodies should be potentially efficacious in treating asthma with an atopic component. Recent evidence supports this view, and several recent studies (36, 40, 47) have indicated that seasonal pollen-induced asthma and house-dust-mite asthma respond to immunotherapy.

Frankland and Augustin (48) have demonstrated that timothy and orchard grass pollen immunotherapy had a beneficial effect on bronchial symptoms during grass pollen season. Several other European studies have demonstrated that grass immunotherapy benefitted patients with bronchial asthma (49, 50). Although these observations indicate that patients who have grass pollen asthma improve with immunotherapy, other studies attempting to demonstrate benefits of immunotherapy for ragweed-induced asthma have not been as conclusive (51, 52). The reason for this lack of benefit is unclear, but some of the patients in these study groups had sensitivities to other aeroallergens, including molds.

Significant benefit has been described from immunotherapy for house dust mite (53–55), including reduced bronchial hyperreactivity (56). New improved and standardized dust mite extracts have significantly contributed.

Data on immunotherapy for mold-induced asthma is not as extensive, but several studies have demonstrated clinical benefit with immunotherapy for *Cladosporium* (57) and *Alternaria* (58). Unlike dust mite and grass pollen immunotherapy, mold immunotherapy poses potential risk because it has been associated with type III allergic reactions (59). Although mold immunotherapy may be effective in asthma, more studies and extracts of better quality are needed before it can be recommended.

Similarly, data regarding immunotherapy for animal-dander-induced asthma is inconclusive, and avoidance of the provoking animal is still preferred over immunotherapy (60, 61).

## Summary

Allergy plays an important role in chronic asthma. Aeroallergens, both indoor and outdoor, provide a constant source of bronchial inflammation that increases bronchial hyperreactivity. This increased reactivity lowers the threshold of response to nonspecific triggers of asthma flares such as viral infections, pollutants, and cold air (62).

Asthma is a chronic inflammatory disease in much the same way that the rheumatic diseases are. In that respect, therapy should be directed at reducing inflammation, as it is in the rheumatic diseases. Along these lines, newer approaches to asthma therapy should emphasize steroids as first-line treatment (63). In this way, therapy is targeted against the underlying inflammation that causes the bronchial obstruction, rather than treating with bronchodilators such as theophyline and B agonists. The effect of such continuous antiinflammatory therapy is decreased airway hyperresponsiveness, which has obvious clinical benefit.

## References

1. Baushey HA, Holtzman MJ, Sheller JR, Nadel JA. Bronchial hyperreactivity. Am Rev Respir Dis 1985; 121:389–414.
2. Busse WW. The role of inflammation in asthma: a new focus. J Rev Respir Dis 1989; 10:72–80.
3. Djukanovic R, Roche WR, Wilson JW, Beasly CRW, Twentyman OP, Howarth PH, Holgate ST. Mucosal inflammation in asthma. Am Rev Respir Dis 1990; 142:434–457.
4. Beasley R, Roche WR, Roberts JA, Holgate ST. Cellular events in the bronchi in mild asthma and after bronchial provocation. Am Rev Respir Dis 1989; 139:808–817.
5. Laitinen LA, Heino M, Laitnen A, Kava A, Hoahleta T. Damage of the airway epithelium and bronchial reactivity in mild asthma and after bronchial reactivity in patients with asthma. Am Rev Respir Dis 1985; 131:599–606.
6. Djukanovic R, Wilson J, Britten K, et al. Quantitation of mast cells and eosinophils in the bronchial mucosa of symptomatic atopic asthmatics and healthy control subjects using immunohistochemistry. Am Rev Respir Dis 1990; 142:863–871.
7. Holgate ST, Beasley R, Twentyman OP. The pathogenesis and significance of bronchial hyperresponsiveness in airways disease. Clin Sci 1987; 73:501–572.
8. Bleecker ER. Airways hyperresponsiveness and asthma: significance and treatment. J Allerg Clin Immun 1985; 75:21–24.
9. Cockroft DW. Airway hyperresponsiveness: therapeutic implications. Ann Alerg 1987; 59:405–414.
10. Baushey HA, Holtzman MJ, Sheller JR, Nadel JA. Bronchial hyperreactivity. Am Rev Respir Dis 1980; 121:389–414.
11. O'Byrne P, Dolovitch J, Hargreave FE. Late asthmatic responses. Am Rev Respir Dis 1987; 134:740–751.
12. Metzer WJ, Zavala D, Richardson HB, et al. Local allergen challenge and bronchial alveolar lavage of allergic asthmatic lungs. Am Rev Respir Dis 1987; 135:433–440.

13. Wardlaw AJ, Dunnett S, Gleich GJ, Collins JV, Kay AB. Eosinophils and mast cells in bronchoalveolar lavage in subjects with mild asthma. Am Rev Respir Dis 1988; 137:62–69.
14. Gibson PG, Girgis-Gabardo A, Morris MM, et al. Cellular characteristics of sputum from patients with asthma and chronic bronchitis. Thorax 1989; 44:689–692.
15. Busse WW, Reed C, Hoehne JH. Where is the allergic reaction in ragweed asthma? II. Demonstration of ragweed antigen in airbourne particles smaller than pollen. J Allerg Clin Immun 1972; 50:289.
16. Solomon WR. Uncovering the "fine details" of pollen allergen transport. J Allerg Clin Immunol 1984; 74:674.
17. Wiken AF, Novey HS, Benhe R, Suprenant EL. Deposition of inhaled pollen and pollen extracts in human airways. N Engl J Med 1973; 288:1056.
18. Gergen PJ, Turkeltaub PC. The association of allergen skin test reactivity and respiratory disease among whites in the US population. Arch Int Med 1991; 151:487–494.
19. Burrows B, Martinez FD, Halonen M, et al. Association of asthma with serum IgE levels and skin test reactivities to allergens. N Engl J Med 1989; 320:271–277.
20. Kalliel JN, Goldstein BM, Broman SS, et al. High frequency of atopic asthma in a pulmonary chest clinic population. Chest 1989; 96:1336–1340.
21. Pollart SM, Chapman M, Fiocco GP, Rose G, Platts-Mills T. Epidemiology of acute asthma: IgE antibodies to major aeroallergens as a risk factor for emergency room visits. J Allerg Clin Immunol 1989; 83:875–82.
22. Licorish K, Novey HS, Kozak P, et al. Role of *Alternia* and *Penicillium* spores in the pathogenesis of asthma. J Allerg Clin Immunol 1985; 76:819–826.
23. Barbee RA. The epidemiology of asthma. Chest (suppl) 1987; 87:217–227.
24. Ellis EF. Asthma in infancy and childhood. In Middleton E, Reed CE, Ellis EF, eds. Allergy: principles and practice, 3rd ed., chap. 45. St. Louis: C.V. Mosby, 1988.
25. Zimmerman B, Feanny L, Reisman J, et al. Allergy in asthma: I. The dose relationship of allergy to severity of childhood asthma. J Allerg Clin Immunol 1988; 81:63–70.
26. Martin AJ, Landau LI, Phelan PD. Natural history of allergy in asthmatic children followed to adult life. Med J Aust 1981; 2:470.
27. Zimmerman B, Chambers C, Forsyth L. Allergy in asthma: II. The highly atopic infant and chronic asthma. J Allerg Clin Immunol 1988; 81:71–77.
28. Maccia CA, Bernstein IL, Emmett EA, et al. In vitro demonstration of specific IgE in phthalmic anhydride hypersensitivity. Am Rev Respir Dis 1976; 113:701–706.
29. Davies RJ, Hendrik DJ, Pepys J. Asthma due to inhaled chemical agents: ampicilin, benzyl penicillin, 6-amino penicillanic acid and related substances. Clin Allerg 1974; 5:99.
30. Bernstein DI, Gallagher JS, D'Sousa L, Bernstein IL. Occupational asthma. J Allerg Clin Immunol 1984; 74:794–803.
31. Inouye T, Tarbo S, Broder I, et al. Severity of asthma in skin test negative and skin test positive patients. J Allerg Clin Immunol 1985; 75:313.
32. Samter M, Beers RF. Intolerance to aspirin. Ann Int Med 1968; 68:975–983.
33. Samuellsson B. Leukotrienes: mediators of immediate hypersensitivity reactions and inflammations. Science 1983; 220:568–570.
34. Murray AB, Ferguson AC, Morrison B. The seasonal variation of allergic respiratory symptoms induced by house dust mite. Ann Allerg 1980; 45:347.
35. Bruce CA, Norman PS, Rosenthal PR, Lichebstein LM. The role of ragweed pollen in autumnal asthma. J Allerg Clin Immunol 1977; 59:444–459.
36. Reid MJ, Moss RB, Hsu YP, et al. Seasonal asthma in northern California: allergic causes and efficiency of immunotherapy. J Allerg Clin Immunol 1986; 78:590–600.
37. Davies RP, Smith LP. Weather and the grass pollen content of the air. Clin Allerg 1973; 3:95.
38. Newmark FN, Itkin IH. Asthma due to pine pollen. Ann Allerg 1967; 25:251.
39. Ogden EC, Raymon GS, Hagen JV, et al. Manual for sampling airborne pollens. New York: Hafner Press, 1954.
40. Ortolani, Pastorello E, Moss RB, et al. Grass pollen immunotherapy: a single year, double blinded placebo controlled study in patients with grass induced asthma. J Allerg Clin Immunol 1984; 73:283–290.
41. Lewis WH, Vinay P, Zenger VE. Airborne and allergic pollen of North America. Baltimore: Johns Hopkins University Press, 1983.
42. Skin testing and radioallergosorbent (RAST) for diagnosis of specific allergens responsible for IgE mediated disease: position statement. J Allerg Clin Immunol 1983; 72:515.
43. Curran WS, Goldman G. The incidence of immediately reacting skin tests in a "normal" adult population. Ann Int Med 1961; 55:777.
44. Hagy GW, Settipane GA. Prognosis of positive allergy skin tests in an asymptomatic population: a three year follow up of college students. J Allerg Clin Immunol 1971; 48:200.
45. Hagy GW, Settipane GA. Risk factors for developing asthma and allergic rhinitis. A 7 year follow up of college students. J Allerg Clin Immunol 1976; 58:330.
46. McAllen MK. Bronchial sensitivity testing in asthma. Thorax 1961; 16:30–37.
47. Bousquet J, Hejiooui, Michael FB. Specific immunotherapy in asthma. 1990; 86:292–305.
48. Frankland AW, Augustin R. Prophylaxis of summer hay fever and asthma: a controlled trial comparing crude grass-pollen extracts with the isolated main protein component. Lancet 1954; 1:1055–1057.
49. Osterballe O. Immunotherapy in hay fever with two major allergens, 19, 25 and partially purified extract of timothy-grass pollen: a controlled double-blind study; in vivo variables, season I. J Allerg 1980; 35:473–489.
50. Johnson DE, Dutton A. The value of hyposensitization therapy for bronchial asthma in children. Pediatrics 1968; 42:793–802.
51. Bousquet J, Guerin B, Dotte A, et al. Comparison of rush immunotherapy with a standardized grass-pollen extract and classical immunotherapy with a pyridine extracted alum adjuved extract. Clin Allerg 1985; 15:179–194.
52. Bruce CA, Norman PS, Lichtenstein LM. The role of ragweed pollen in autumnal asthma. J Allerg Clin Immunol 1977; 59:449–459.
53. Bousquet J, Hejjaoui A, Clauzel AM, et al. Specific immunotherapy with a standardized *Dermatophagoiodes pteronyssinus* extract: II. Prediction of efficacy of immunotherapy. J Allerg Clin Immunol 1988; 82:991–997.
54. D'Souza ME, Pepys J, Wells ID, et al. Hyposensitization with *Dermatophagoides pteronyssinus* in house dust allergy: a controlled study of clinical and immunological effects. Clin Allerg 1973; 3:177–193.

55. Formgren H, Lanner A, Linholm N, Lowhagen O, Dreborg S. Effects of immunotherapy on specific and non-specific sensitivities of the airways. J Allerg Clin Immunol 1984; 73:140.
56. Bousquet J, Calvayrac P, Guerin B, et al. Immunotherapy with a standardized *Dermatophagoides pteronyssinus* extract: I. In vivo and in vitro parameters after a short course of treatment. J Allerg Clin Immunol 1985; 76:734–744.
57. Karlksson R, Agrell B, Dreborg S, et al. A double-blinded multicentered immunotherapy trial in children using a purified and standardized *Cladosporium herbarum*. Allergy 1986; 41:140–147.
58. Horst M, Hejjaoui A, Horst V, Michel F-B, Bousquet J. Double blinded placebo controlled rush immunotherapy with a standardized *Alternia* extract. J Allerg Clin Immunol 1990; 85:460–472.
59. Busse WW, Storms WW, Flaherty DK, Crandell M, Reed CE. *Alternia* IgG precipitins and adverse reactions. J Allerg Clin Immunol 1976; 57:367–372.
60. Lilja G, Sundlin B, Graff-Lonnevig V, et al. Immunotherapy with cat and dog dander extracts: IV. Effects of 2 years of treatment. J Allerg Clin Immunol 1989; 83:37–44.
61. Sundlin B, Lilja G, Graff-Lonnevig V, et al. Immunotherapy with partially purified and standardized animal extracts: I. Clinical results from a double-blinded study on patients with animal dander asthma. J Allerg Clin Immunol 1986; 77:478–487.
62. Chung KF. Role of inflammation in the hyperreactivity of the airways in asthma. Thorax 1986; 41:657–662.
63. Konig P. Inhaled corticosteroids—the present and future role in the management of asthma. J Allerg Clin Immunol 1988; 82:297–306.

# Asthma in Children and Its Management:

## An Update

DIANA LOWENTHAL, M.D., AND MEYER KATTAN, M.D.

ASTHMA is a common cause of morbidity and school absence in childhood. Recent studies point to an increasing incidence and prevalence of childhood asthma and a rise in hospitalization and mortality rates in the United States and elsewhere (1–5). Gergen et al (1) reported that the percentage of children aged 6–11 years with asthma during the previous year increased from 4.8% during the years 1971–1974 to 7.6% during the years 1976–1980. This report also found a higher prevalence in blacks and urban residents. Asthma mortality among persons aged 5 to 34 years in the United States declined during the 1970s, but increased 6.2% yearly during the 1980s (6). Mortality rates are higher among urban residents and blacks (6, 7).

The reasons for the increased morbidity and mortality from asthma are unknown. Suggested explanations include increasing exposure to air pollution and other environmental triggers, inadequate access to health care, patient and physician underestimation of the severity of asthma, and suboptimal medical management. Some of the reported increase may be a result of improved diagnosis, because the assigned diagnosis of acute or chronic bronchitis, an entity which does not appear to occur in childhood, has been declining since 1983 (2).

Despite the importance of asthma as a cause of childhood illness, the condition remains both underdiagnosed and undertreated. Speight et al (8) surveyed the parents of third graders and found that only 35% of children having more than 12 episodes of wheezing per year were diagnosed with asthma. No bronchodilator had ever been tried as treatment for two-thirds of the children with a history of wheezing. The diagnoses described to the parents of the children who had asthma included bronchitis, allergic chest condition, and recurrent viruses, and many children received antibiotics regularly for their respiratory problems. The authors documented a tenfold decrease in school absenteeism in the children having over 12 episodes of wheezing per year after they were treated with regular asthma medication. Rather than being apprehensive, the parents were grateful for a diagnosis and an explanation of their child's respiratory problems.

Although the classic initial symptom of asthma is wheezing, several other patterns occur; some children have nocturnal or morning cough, dyspnea, or chest tightness. Cloutier reported chronic cough as a manifestation of hyperreactive airways (9). Exercise-induced wheeze or chest tightness is another common manifestation of asthma.

Several studies have suggested a link between viral infection and the onset of asthma. Welliver (10) found that children with respiratory syncytial virus (RSV) with wheezing had higher acute and convalescent titers of RSV-specific IgE than children who had RSV upper respiratory tract illness or pneumonia without wheeze. Similar findings were documented in children with parainfluenza virus resulting in croup or wheezing (11). Children with RSV infection and high RSV-IgE had a 70% incidence of wheezing within the next 48 months, whereas only 28% of those

---

From the Pediatric Pulmonary & Critical Care Division, Department of Pediatrics, Mount Sinai School of Medicine, One Gustave L. Levy Place, New York, NY. Address correspondence and reprint requests to Meyer Kattan, M.D., Pediatric Pulmonary and Critical Care Division, Box 1202, Mount Sinai School of Medicine, One Gustave L. Levy Place, New York, NY 10029.

with undetectable RSV-IgE had subsequent wheezing (12). A review of the work linking viral infection and the onset of asthma has recently been published (13).

## Principles of Management

The goal of management is to make the child symptom-free. This allows the child to attend school, participate in sports, and sleep comfortably. It also avoids family disruption and lost work days for parents. A treatment regimen must be custom-tailored to each patient's needs. Some patients may need medication intermittently, or only with exercise or viral infections. Others require daily treatment. It is important to consider the lifestyle of the family. For example, many families are unable to comply with a regimen which requires that medication be taken every six hours. A nebulizer may be inconvenient for a family which spends much time outdoors or away from home.

The patient and family must be able to recognize the early signs of an asthma attack, as well as signs of respiratory distress. There are individual differences in perception of airway obstruction. Some patients feel comfortable in the face of severe airway obstruction. These patients must be treated on the basis of objective data, such as peak expiratory flow rate (PEFR) or spirometry.

Physicians must actively encourage families to avoid where possible asthma triggers, which may include cigarette smoke, animals, or house dust. Many asthmatics are sensitized to house dust mites and benefit from zippered bedding covers and wet-mopping floors (14, 15). Immunotherapy may be of benefit in selected patients.

It is crucial to consider psychosocial issues in the management of asthma. Psychiatric problems may trigger attacks, and some patients or their families may use asthma in manipulative ways. One study found that the risk of dying from asthma was associated with several psychological problems, including emotional disturbance, depressive symptoms, disregard of asthma symptoms, manipulative use of asthma, and family dysfunction (16). Anxiety disorders appear to be increased in asthmatics, and may result in a cycle whereby asthma triggers a panic attack, which results in worsening asthma (17). Vocal cord dysfunction, in which there is paradoxical vocal cord motion with respiration, is a functional disorder often diagnosed as asthma. These patients improve with a combination of psychiatric consultation and behavioral modification (18).

### Pharmacotherapy

The goal of medical management is to treat both the bronchoconstriction and the inflammation. Several agents are available, some more useful for acute episodes. Specific management guidelines have recently been published (18a).

**Inhalation.** Inhalation is an effective method of delivery of medications because it allows for deposition of medication to the lungs with minimal systemic absorption.

Three inhalation methods are available: inhalation of powder through a Spinhaler device, inhalation with a metered-dose inhaler, and inhalation of an aerosolized solution via a nebulizer. The nebulizer cup contains medication, through which compressed air or oxygen is forced, forming an aerosol. The child can breathe this aerosol through either a facemask or a mouthpiece with equal efficacy (19). Using a metered-dose inhaler requires a degree of coordination which is difficult for some children and adults. It is important to coordinate activation of the inhaler with each breath, to inhale slowly, and to breathhold for several seconds after inhalation.

To facilitate metered-dose inhaler use, a variety of "spacers" and "holding chambers" have been devised. The patient sprays the inhaler medication into the spacer, often a cylindrical tube or bag. The patient then inhales from the mouthpiece of the spacer. Thus, the need to coordinate spraying and inhaling is eliminated. Some spacers have an audible or other cue which encourages a slow inhalation. Children as young as four have been taught to use an inhaler with a spacer attachment. A review of spacers has been published by Levison et al (20).

**β-Adrenergic Agonists.** Adrenergic bronchodilators have been widely used to treat acute asthma. There has been a shift to inhaled, instead of injected, agents, and the use of more $\beta_2$-selective agents in acute asthma. Inhaled agents are at least as effective as injected agents even in severely obstructed patients, with fewer side effects (21). Epinephrine can be administered acutely. However, the commonly used regimen of repeated injections offers little advantage over a single injection of Sus-Phrine (long-acting epinephrine) (22).

The most commonly used inhaled beta-agonists in use in the United States are albuterol and metaproterenol. (Table). Inhaled beta-agonists have an onset of action within minutes. They are typically used either as needed or up to every six hours by stable outpatients. The inhaled beta-agonists can be used more frequently or even continually (by nebulizer), but such use should be in

TABLE

*Available Forms and Dosage of Inhaled Adrenergic Agents*

| | Powder (inhalations) | Metered-dose inhaler (puffs) | Nebulizer (mL/kg) |
|---|---|---|---|
| Albuterol | 2 | 2 | 0.01–0.03<br>(max 1.0 mL)<br>of 0.5% solution |
| Metaproterenol | — | 2 | 0.01<br>(max 0.3 mL)<br>of 5% solution |

the hospital to monitor the patient. Emergency room studies have found greater bronchodilation with high-dose (0.15 mg/kg) than with low-dose (0.05 mg/kg) nebulized albuterol (23). Treatments given every 20 minutes were more effective than higher doses every hour (24). Common adverse reactions to inhaled albuterol are tremor, which is generally mild and transient, and tachycardia.

Beta-agonists are also available in oral form, which is useful for children who are too young to use metered-dose inhalers and those who prefer oral medication.

Since $\beta_2$-agonist use is effective and well tolerated, recent trends have encouraged regular beta-agonist use for chronic asthmatics. However, this concept may need reevaluation. A recent double-blind, randomized crossover study found asthma control to be better in patients who took beta-agonists only as needed for symptomatic relief, compared to those taking regular daily doses (25). In addition to use in acute asthma, beta-agonists are also effective prophylactic agents for exercise-induced asthma.

**Anticholinergic Agents.** Vagal stimulation causes bronchoconstriction and increased mucus production. Recent studies suggest a role for anticholinergic agents, such as ipratropium bromide, a derivative of atropine, in acute asthma. Two studies (26, 27) found improved bronchodilation when inhaled ipratropium and albuterol were administered for an acute asthma attack in children, compared to albuterol alone.

**Theophylline.** Theophylline is a commonly used bronchodilator. Sustained-release products allow for convenient dosing. Its bronchodilator effect is directly proportional to serum level. Unfortunately, adverse effects, including nervousness and gastrointestinal upset, are common even in the therapeutic serum range of 10–20 $\mu$g/mL, and often limit patient compliance. In the toxic range, seizures, arrhythmia, and death may occur. Another disadvantage to theophylline is that serum levels are affected by numerous factors, including time of day (28), food intake (29), smoking (30), acute viral infection (31), and many medications, including erythromycin (32) and cimetidine (33).

Theophylline offers no additional therapeutic benefit to inhaled beta-agonist in the initial management of an acute attack (34). A recent study of hospitalized asthmatics found no benefit to adding theophylline to a regimen of frequent nebulized albuterol and oral prednisone (35). Although the efficacy and safety of new bronchodilators and prophylactic agents have diminished the role for theophylline, it is still useful for treatment of chronic asthma in certain patients. It is available in numerous preparations, including rapid- and slow-release oral preparations and intravenous preparations.

**Disodium Cromoglycate (Cromolyn).** Disodium cromoglycate appears to stabilize the mast cell membrane, preventing degranulation in response to antigen. The efficacy of this medication as a prophylactic agent when used daily has been demonstrated (36). In addition, it may be used prior to exercise, to prevent exercise-induced asthma (37). It can be inhaled in a metered-dose inhaler or nebulizer, but no oral form is available. It is virtually free of significant adverse effects. Its efficacy and lack of side effects make it a preferred medication for those children requiring daily therapy.

**Corticosteroids.** In recent years, corticosteroids have received increasing attention and use for the treatment of asthma. These agents have anti-inflammatory properties that may reduce edema and mucus production. Studies have demonstrated that steroids decrease bronchial hyperresponsiveness (38–40) and prevent the late asthmatic response to antigen challenge (40). Steroids improve pulmonary function in chronic asthma. A dose of corticosteroid produces greater pulmonary function improvement than placebo when bronchodilators are withheld in patients with stable asthma (41, 42). Steroids also increase responsiveness to a beta-agonist in the stable asthmatic with a poor response to bronchodilator (43).

The role of steroids for the management of an acute asthmatic exacerbation is less certain. Some studies have demonstrated that children with an exacerbation of asthma who were treated with steroids had a better clinical response and were less likely to require a steroid "rescue" than patients treated with placebo (44, 45). Another showed that children treated with steroids for an exacerbation had less wheezing and better pulmonary function after seven days than their placebo-treated counterparts (46). Storr showed that children over the age of two had a shorter hospitalization if treated with steroids initially (47). Pierson found that a steroid-treated group of hospitalized children had a higher $paO_2$ at 3 and 24 hours, although pulmonary function results and clinical scoring did not differ between the steroid- and placebo-treated groups (48).

Littenberg, in a study of adults with an acute exacerbation, found that the steroid-treated group had a lower admission rate, and patients reported better subjectively measured functioning than the placebo group, although objective pulmonary function tests were not different between the two groups (49). Fanta studied adult asthmatics who were unresponsive to four hours of frequent bronchodilator therapy, and then randomized them to receive either steroids or placebo. The steroid-treated group had higher percentage improvement in $FEV_1$ at 24 hours, but neither group had an improvement in $paO_2$ at 24 hours (50).

Other studies have not shown any benefit to administering steroids for acute asthma management. Stein, in a recent study of 81 adult asthmatics, demonstrated that steroid use in the emergency room had no effect on pulmonary function, duration of emergency room treatment, hospitalization rate, or revisit rate in patients discharged (51). A study of children hospitalized with asthma, all of whom received aminophylline and albuterol by nebulizer, found no difference in clinical score or PEFR between those who received intravenous steroids and those who received placebo (52). McFadden studied 38 adult asthmatics in an emergency room setting and randomized them to receive a single dose of intravenous hydrocortisone (either 0.25, 0.50, or 1.0 g) or placebo. The patients were then studied for at least six hours. No benefit of steroid treatment, at any dose, over placebo could be demonstrated at any point in time, based on plethysmography, spirometry, or number of beta-agonist treatments required (53).

Overall, studies of steroids for acute asthma have reported mixed results. Many of these studies are difficult to interpret because of small sample size, inadequately controlled design, steroid use in the control group, or underutilization of beta-agonist. A critical review of these issues was published in 1985 (54). Additional large, well-controlled studies of steroids for acute asthma are needed. Although some subgroups of asthmatics may benefit from steroids, review of the current literature does not fully support steroid use for all asthmatic exacerbations. The risk of treating all exacerbations with steroids, or prophylactically administering steroids for all respiratory infections, is that children may end up with numerous and frequent courses of corticosteroids, which may result in adrenal insufficiency.

**Inhaled Steroids.** A recent trend has favored the use of inhaled over oral steroids. Studies have demonstrated that inhaled steroids diminish bronchial reactivity (38, 55) and limit the maximal degree of airway narrowing (56). High-dose inhaled steroids maintain the pulmonary function improvement induced by a course of oral steroids (57).

Some studies found no adrenal suppression when inhaled corticosteroids were used (58, 59). Many of these studies relied on random or morning cortisol levels or tests of adrenocorticotropic hormone (ACTH) stimulation. However, another study of patients using inhaled corticosteroids found abnormal cortisol response to insulin-induced hypoglycemia in five patients with normal response to cosyntropin stimulation, suggesting that response to hypoglycemia may be a more sensitive test for adrenal insufficiency (60). A study comparing inhaled beclomethasone with alternate-day oral prednisone or a combination found that morning cortisol concentration, urine-free cortisol excretion, and 11-deoxycortisol response to metyrapone were decreased to a similar degree with either regimen, and were lowest with the combination (61). A recent study measured cortisol levels every 30 minutes for a 24-hour period in 10 children with asthma using inhaled steroids (62). This study found much lower levels of cortisol secretion, especially at night, in the asthmatics than in matched controls.

Although the clinical signs of adrenal insufficiency have not been documented in asthmatics using inhaled steroids, these agents should be used with caution in the growing child until long-term studies of effects in this population are available.

**Compliance.** Once a medical regimen is prescribed, one must continuously evaluate patients for compliance. Erratic compliance with prophylactic medications may render the medication practically useless. One must also regularly evaluate the inhalation technique of patients using metered-dose inhalers.

## Outcome of Childhood Asthma

It is widely believed that children tend to outgrow asthma. However, long-term studies of asthmatics do not support this view.

Ryssing (63) surveyed 281 patients who had been hospitalized with asthma 18 to 31 years earlier and found that only 28% were free of asthmatic symptoms. Bronnimann (64) surveyed 2300 asthmatics of all ages on two occasions, with a mean follow-up of 9.4 years. By the second survey, 22% of the asthmatics were in remission. The highest remission rate, 65%, was found in patients aged 10–19 years. However, relapses were common at all ages, and increased with age. For example, the relapse rate was 30% for patients in their 20s and over 60% for those in their 60s. The initial study identified 99 patients as ex-asthmatics. By the second study, 38% had relapsed.

Kelly (65) found that severity of wheezing at age 14 years was similar to that at 28, and that 44% of those with infrequent wheezing at age 21 years had deteriorated by age 28. Blair (66) followed up 244 children for 20 years and found that 52% were nearly or completely free of symptoms after 20 years. The final prognosis was influenced by initial asthma severity, history of atopy in the patient or first-degree relatives, and breastfeeding. Gerritsen (67) reevaluated asthmatic children after a 16-year interval, and found that 43% had current symptoms and 27% were taking daily medications. Although the degree of airway obstruction as measured by pulmonary function testing did not change, the percentage of patients with airway responsiveness to histamine decreased markedly.

Although these studies reported large differences in rates of persistent asthma, probably because of different patient selection criteria and definitions, it seems clear that one cannot expect children to outgrow their asthma, especially if the asthma was not mild in childhood. In addition, many patients who appear to outgrow asthma during adolescence will have relapses as adults.

## References

1. Gergen PG, Mullally DI, Evans R III. National survey of prevalence of asthma among children in the United States, 1976–1980. Pediatrics 1988; 81:1–7.
2. Gergen PG, Weiss KB. Changing patterns of asthma hospitalization among children: 1979–1987. JAMA 1990; 264:1688–1692.
3. Halfon N, Newacheck PW. Trends in the hospitalization for acute childhood asthma, 1970–84. Am J Publ Health 1986; 76:1308–1311.
4. Evans R III, Mullally DI, Wilson RW, et al. National trends in the morbidity and mortality of asthma in the US: prevalence, hospitalization and death from asthma over two decades: 1965–1984. Chest 1987; 91:655–745.
5. Mitchell GA. International trends in hospital admission rates for asthma. Arch Dis Child 1985; 60:376–378.
6. Weiss KB, Wegener DK. Changing patterns of asthma mortality: identifying target population at high risk. JAMA 1990; 264:1683–1687.
7. Malveaux FJ. Asthma in the black community. In: Wissow LS, ed. Pediatric asthma, report of the twenty-first Ross roundtable on critical approaches to common pediatric problems. Columbus, OH: Ross Labs, 1990:3–4.
8. Speight ANP, Lee PA, Hey EN. Underdiagnosis and undertreatment of asthma in childhood. Br Med J 1983; 286:1253–1256.
9. Cloutier MM, Loughlin GM. Chronic cough in children: a manifestation of airway hyperreactivity. Pediatr 1981; 67:6–12.
10. Welliver RC, Wong DT, Sun M, Middleton E Jr, Vaughan RS, Ogra PL. The development of respiratory syncytial virus-specific IgE and the release of histamine in nasopharyngeal secretions after infection. N Engl J Med 1981; 305:841–846.
11. Welliver RC, Wong DT, Middleton E Jr, Sun M, McCarthy N, Ogra PL. Role of parainfluenza virus-specific IgE in pathogenesis of croup and wheezing subsequent to infection. J Pediatr 1982; 101:889–896.
12. Welliver RC, Sun M, Rinaldo D, Ogra PL. Predictive value of respiratory syncytial virus-specific IgE responses for recurrent wheezing following bronchiolitis. J Pediatr 1986; 109:776–880.
13. Busse WW. The relationship between viral infections and onset of allergic diseases and asthma. Clin Exp Allerg 1989; 19:1–9.
14. Murray AB, Ferguson AC. Dust-free bedrooms in the treatment of asthmatic children with house dust mite allergy: a controlled trial. Pediatrics 1983; 71:418–422.
15. Platts-Mills TAE, deWeck AL, et al. Dust mite allergens and asthma—a worldwide problem. J Allerg Clin Immunol 1989; 83:416–427.
16. Strunk RC, Mrazek DA, Wolfson Furhrmann GS, LaBrecque JF. Physiologic and psychological characteristics associated with deaths due to asthma in childhood: a case-controlled study. JAMA 1985; 254:1193–1198.
17. Yellowlees PM, Kalucy RS. Psychobiological aspects of asthma and the consequent research implications. Chest 1990; 97:628–634.
18. Christopher KL, Wood RP, Eckert C, et al. Vocal-cord dysfunction presenting as asthma. N Engl J Med 1983; 308:1566–1570.

18a. Executive Summary: Guidelines for the diagnosis and management of asthma. U.S. Dept of Health and Human Services, pub. no. 91-3042A, June 1991.

19. Lowenthal DB, Kattan M. Mouthpieces vs. facemasks for aerosol treatment of asthmatic children. Am Rev Respir Dis 1991; 143:A21.
20. Levison H, Reilly PA, Worsley GH. Spacing devices and metered-dose inhalers in childhood asthma. J Pediatr 1985; 107:662–668.
21. Ben-Zvi Z, Lam C, Hoffman J, Teets-Grimm KC, Kattan M. An evaluation of the initial treatment of acute asthma. Pediatrics 1982; 70:348–353.
22. Ben-Zvi Z, Lam C, Spohn WA, et al. An evaluation of repeated injections of epinephrine for the initial treatment of acute asthma. Am Rev Respir Dis 1983; 127:101–105.
23. Schuh S, Parkin P, Rajan A, Canny G, et al. High- versus low-dose, frequently administered nebulized albuterol in children with severe, acute asthma. Pediatrics 1989; 83:513–518.
24. Robertson CF, Smith F, Beck R, Levison H. Response to frequent low doses of nebulized salbutamol in acute asthma. J Pediatr 1985; 106:672–674.
25. Sears MR, Taylor DR, Print CB, et al. Regular inhaled beta-agonist treatment in bronchial asthma. Lancet 1990; 336:1391–1396.
26. Beck R, Robertson C, Galdes-Sebaldt M, Levison H. Clinical and laboratory observations: combined salbutamol and ipratropium bromide by inhalation in the treat-

ment of severe acute asthma. J Pediatr 1985; 107:605–608.

27. Reisman J, Galdes-Sebalt M, Kazim F, Canny G, Levison H. Frequent administration by inhalation of salbutamol and ipratropium bromide in the initial management of severe acute asthma in children. J Allerg Clin Immunol 1988; 81:16–20.
28. Scott PH, Tabachnik E, MacLeod S, et al. Sustained-release theophylline for childhood asthma: evidence for circadian variation of theophylline pharmacokinetics. J Pediatr 1981; 99:476–479.
29. Hendeles L, Weinberger M, Milavetz G, et al. Food induced "dose dumping" from a once a day theophylline product as a cause of theophylline toxicity. Chest 1985; 87:758–765.
30. Jenne J, Nagasawa H, McHugh R, et al. Decreased theophylline half-life in cigarette smokers. Life Sci 1975; 17:195–198.
31. Chang KC, Bell TD, Lauer BA, Chai H. Altered theophylline pharmacokinetics during acute respiratory viral illness. Lancet 1978; 1:1132–1133.
32. Renton KW, Gray JD, Hung OR. Depression of theophylline elimination by erythromycin. Clin Pharmacol Ther 1981; 30:422–426.
33. Jackson JE, Powell JR, Wandell M, et al. Cimetidine decreases theophylline clearance. Am Rev Respir Dis 1981; 123:615.
34. Siegel D, Sheppard D, Gelb A, Weinberg PF. Aminophylline increases the toxicity but not the efficacy of an inhaled beta-adrenergic agonist in the treatment of acute exacerbation of asthma. Am Rev Respir Dis 1985; 132:283–286.
35. Self TH, Abou-Shala N, Burns R, et al. Inhaled albuterol and oral prednisone therapy in hospitalized adult asthmatics: does aminophylline add any benefit? Chest 1990; 98:1317–1321.
36. Shapiro GG, Furukawa CT, Pierson WE, et al. Double-blind evaluation of nebulized cromolyn, terbutaline, and the combination for childhood asthma. J Allerg Clin Immunol 1988; 81:449–454.
37. Debelic M, Hertel G, Konig J. Double-blind crossover study comparing sodium cromoglycate, reproterol, reproterol plus sodium cromoglycate, and placebo in exercise-induced asthma. Ann Allerg 1988; 61:25–29.
38. Dutoit JI, Salome CM, Woolcock AJ. Inhaled corticosteroids reduce the severity of bronchial hyperresponsiveness in asthma but oral theophylline does not. Am Rev Respir Dis 1987; 136:1174–1178.
39. Bhagat RG, Grunstein MM. Effect of corticosteroids on bronchial responsiveness to methacholine in asthmatic children. Am Rev Respir Dis 1985; 131:902–906.
40. Fabbri LM, Chiesura-Corona P, Dal Vecchio L, et al. Prednisone inhibits late asthmatic reactions and the associated increase in airway responsiveness induced by toluene-diisocyanate in sensitized subjects. Am Rev Respir Dis 1985; 132:1010–1014.
41. Ellul-Michallef R, Fenech FE. Intravenous prednisolone in chronic bronchial asthma. Thorax 1975; 30:312–315.
42. Klaustermeyer WB, Hale FC. The physiologic effect of an intravenous glucocorticoid in bronchial asthma. Ann Allerg 1976; 37:80–86.
43. Ellul-Micallef R, Fenech FE. Effect of intravenous prednisolone in asthmatics with diminished adrenergic responsiveness. Lancet 1975; 2:1269–1271.
44. Loren ML, Chai H, Leung P, Rohr C, Brenner AM. Corticosteroids in the treatment of acute exacerbation of asthma. Ann Allergy 1980; 45:67–71.
45. Harris JB, Weinberger MM, Nassif E, et al. Early intervention with short courses of prednisone to prevent progression of asthma in ambulatory patients incompletely responsive to bronchodilators. J Pediatr 1987; 110:627–633.
46. Shapiro GG, Furukawa CT, Pierson WE, Gardinier R, Bierman CW. Double-blind evaluation of methylprednisolone versus placebo for acute asthma episodes. Pediatrics 1983; 71:510–514.
47. Storr J, Barrell E, Barry W, Lenney W, Hatcher G. Effect of a single dose of prednisolone in acute childhood asthma. Lancet 1987; 1:879–882.
48. Pearson WE, Bierman CW, Kelley VC, A double-blind trial of corticosteroid therapy in status asthmaticus. Pediatrics 1974; 54:282–288.
49. Littenberg B, Gluck EH. A controlled trial of methylprednisolone in the emergency treatment of acute asthma. N Engl J Med 1986; 314:150–152.
50. Fanta CH, Rossing TH, McFadden ER. Glucocorticoids in acute asthma. Am J Med 1983; 74:845–851.
51. Stein LM, Cole RP. Early administration of corticosteroids in emergency room treatment of acute asthma. Ann Int Med 1990; 112:822–827.
52. Kattan M, Gurwitz D, Levison H. Corticosteroids in status asthmaticus. J Pediatr 1980; 96:596–599.
53. McFadden ER, Kiser R, deGroot WJ, Holmes B, et al. A controlled study of the effects of single doses of hydrocortisone on the resolution of acute attacks of asthma. Am J Med 1976; 60:52–59.
54. Mok J, Kattan M, Levison H. A case against the use of corticosteroids in acute, severe asthma. Pharmacotherapy 1985; 5:331–335.
55. Kerrebijn KF, van Essen-Andvliet EEM, Neijens HJ. Effect of long-term treatment with inhaled corticosteroids and beta-agonists on the bronchial responsiveness in children with asthma. J Allerg Clin Immunol 1987; 79:653–659.
56. Bel EH, Timmers MC, Zwinderman AH, Dijkman JH, Stark PJ. The effect of inhaled corticosteroids on the maximal degree of airway narrowing to methacholine in asthmatic subjects. Am Rev Respir Dis 1991; 143:109–113.
57. Salmeson S, Guerin SC, Gonard P, et al. High doses of inhaled corticosteroids in unstable chronic asthma: a multicenter double-blind, placebo-controlled study. Am Rev Respir Dis 1989; 140:167–171.
58. Goldstein D, Konig P. Effect of inhaled beclomethasone dipropionate on hypothalamopituitary adrenal axis function in children with asthma. Pediatrics 1983; 72:60.
59. Klein R, Waldman D, Kershner H, et al. Treatment of chronic childhood asthma with beclomethasone dipropionate aerosol: I. A double-blind, crossover trial in non-steroid-dependent patients. Pediatrics 1977; 60:7–13.
60. Vaz R, Senior B, Morris M, Binkiewitz A. Adrenal effects of beclomethasone inhalation therapy in asthmatic children. J Pediatr 1982; 100:660–662.
61. Wyatt R, Waschek J, Weinberger M, Sherman B. Effects of inhaled beclomethasone dipropionate and alternate-day prednisone on pituitary-adrenal function in children with chronic asthma. N Engl J Med 1978; 299:1387–1392.

62. Tabachnik E, Zudik Z. Diurnal cortisol secretion during therapy with inhaled beclomethasone dipropionate in children with asthma. J Pediatr 1991; 118:294–297.
63. Ryssing E. Continued follow-up investigation concerning the fate of 298 asthmatic children. ACTA Paed 1959; 48:255–260.
64. Bronnimann S, Burrows B. A prospective study of the natural history of asthma. Chest 1986; 90:484–490.
65. Kelly WJW, Hudson I, Raven J, et al. Childhood asthma and adult lung function. Am Rev Respir Dis 1988; 138:26–30.
66. Blair H. Natural history of childhood asthma. Arch Dis Child 1977; 52:613–619.
67. Gerritsen J, Koeter GH, Postma DS, Schauten JP, Knol K. Prognosis of asthma from childhood to adulthood. Am Rev Respir Dis 1989; 140:1325–1330.

# Pulmonary Function Testing in Bronchial Asthma:

## Standard and Emerging Techniques

Lee K. Brown, M.D.

THE DEFINITION of bronchial asthma has long been a subject of controversy in the medical literature. The most recent American Thoracic Society definition is "a clinical syndrome characterized by increased responsiveness of the tracheobronchial tree to a variety of stimuli" (1). This increased responsiveness is manifest by a variety of pathologic airway changes which are much less controversial. These range from smooth muscle contraction and mucus hypersecretion during acute exacerbations, to smooth muscle hypertrophy and hyperplasia, submucosal and mucosal edema, and glandular hypertrophy and hyperplasia (2), which may persist even when the patient is asymptomatic. All of these processes reduce airway cross-sectional area and increase resistance to airflow.

During an acute exacerbation, obstruction to airflow typically occurs in both large and small airways (small airways are conventionally defined as those below 2 mm in diameter), reflected in alteration of pulmonary function tests of large airway function. The pathologic changes of asthma in remission, on the other hand, usually result in significant obstruction only in the small airways, where more sensitive tests are necessary to detect dysfunction (3). It should be stressed that, even in fatal asthma, pathologic changes in the lung parenchyma are absent. Therefore, parameters of pulmonary function associated with parenchymal dysfunction are generally normal (4).

### Spirometric Tests of Pulmonary Function

John Hutchinson perfected the water-sealed spirometer in 1846, and used it extensively for the measurement of vital capacity (VC). However, it was not until the work of Tiffeneau (5) and Gaensler (6), starting in 1947, that spirometric assessment of airflow (rather than volume) became available.

Their work was based on the forced expiratory spirogram: measuring volume vs. time while the subject exhales with maximal effort from total lung capacity (TLC) to residual volume (RV). By measuring the volume exhaled after a fixed time interval from the start of the effort, a parameter equivalent to an average flow could be obtained (volume/time). These timed volumes were found to be reduced when normal subjects exhaled through narrow tubes simulating airways obstruction. In addition, the extent of the reduction in timed volume correlated with the severity of the obstruction (6).

Historically, a variety of time intervals have been utilized in this measurement. Forced expiratory volume in 1 second ($FEV_1$) is now most commonly employed in this regard; with one exception, reduced $FEV_1$ signals obstructive impairment. The exception occurs when restrictive disease (reduced VC) is present. $FEV_1$ may then be reduced roughly in proportion to VC simply because there is less volume available for exhalation, rather than because of an obstructive limitation to airflow. The parameter $FEV_1/FVC$ (where FVC is forced VC-measured during the forced expiratory spirogram) will be normal or elevated in restrictive disease and reduced in obstructive diseases such as asthma (7).

$FEV_1$ and $FEV_1/FVC$ are measurements of

---

From the Division of Pulmonary and Critical Care Medicine, Mount Sinai School of Medicine, New York, NY. Address correspondence and requests for reprints to Lee K. Brown, M.D., Division of Pulmonary and Critical Care Medicine, Box 1232, Mount Sinai Medical Center, New York, NY 10029.

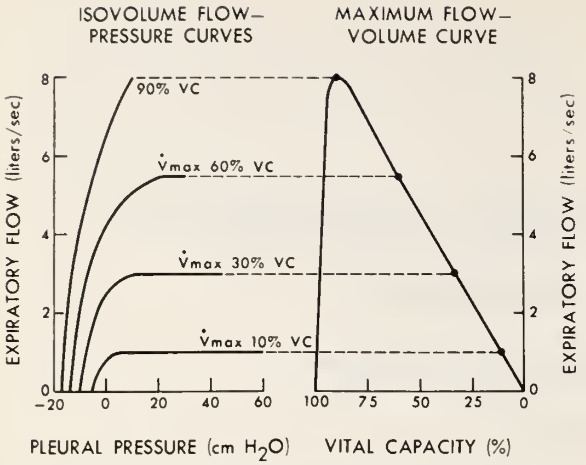


Fig. 1. Effort-independence of expiratory airflow at low lung volume. A family of isovolume pressure-flow curves (left) is related to the maximal expiratory flow-volume curve (right). Note that, at 90% VC, increasing pleural pressure results in increased flow. At 60% VC and below, flow will not increase further once pleural pressure exceeds a threshold value. *Reproduced* with permission from Murray JF. The normal lung: the basis for diagnosis and treatment of pulmonary disease. Philadelphia: WB Saunders, 1976:102.

relatively high degrees of airways obstruction, said to reside in the large airways (greater than 2 mm in diameter). There exist other spirometric tests of small airways obstruction (also termed minimal airways dysfunction), which depend upon the concept of dynamic compression and the "equal pressure point" (8). During a forceful exhalation, "upstream" pressure (in the alveoli) exceeds the positive pleural driving pressure due to the pressure developed by lung elastic recoil. This pressure differential is maintained proceeding "downstream" in the conducting airways toward the mouth, until the pressure drop caused by airways resistance results in an airway pressure equal to pleural pressure. At this point (the "equal pressure point") any increase in expiratory force (pleural pressure) will compress and occlude the airway. When this occurs, airflow ceases, airway pressure rises, the airway is forced open, and the process repeats. In practice, the airway probably does not open and close, but reaches some equilibrium position such that an increase in expiratory effort does not result in increased flow. This situation, termed "effort independence," results in a flow rate which is inversely proportional to upstream airways resistance (residing in the "small" airways) and proportional to the pressure drop from alveolus to the equal pressure point (equivalent to static recoil pressure, $P_{st}$):

$$\dot{V}_{max} = P_{st}/R_{us} \qquad (1)$$

Effort independence does not hold true throughout the forced expiratory effort (8). At high lung volumes, an isovolume plot of expiratory flow vs. pleural pressure will demonstrate increased flow as pleural pressure increases (Fig. 1). Effort independence may only be demonstrated at lung volumes less than approximately 70% of TLC. Thus, the flows at these lower lung volumes may be used as indicators of minimal airways dysfunction, such as bronchial asthma in remission. The forced expiratory flow between 25% and 75% of FVC ($FEF_{25-75\%}$; formerly called maximal midexpiratory flow rate or MMEF) was first proposed by Leuallen and Fowler as such an index (9). It is an average flow, calculated as the slope of a line drawn between the 25% and 75% points on the forced expiratory spirogram, and correlates well with the presence of airways obstruction. In addition, $FEF_{25-75\%}$ is measured largely during the effort-independent portion of the spirogram, and should therefore reflect small airways function. This hypothesis has been verified in a number of studies comparing preoperative spirometry with histopathologic findings in smokers undergoing lung resection for peripheral malignancies (10, 11).

$FEF_{25-75\%}$ will be reduced in restrictive lung disease in the same manner as $FEV_1$. The parameter mid-expiratory time (MET), also known as the forced expiratory time between 25% and 75% of FVC ($FET_{25-75\%}$), was developed to circumvent this problem. Mathematically, $FET_{25-75\%}$ is related to $FEF_{25-75\%}$ by the equation

$$FET_{25-75\%} = FVC/2\ FEF_{25-75\%} \qquad (2)$$

The presence of the FVC term in the numerator normalizes the value of $FET_{25-75\%}$ for changes in FVC (if $FEF_{25-75\%}$ is reduced due to restriction, FVC will be reduced to a similar degree), and prolonged $FET_{25-75\%}$ is a useful indicator of obstructive impairment even in the presence of restrictive disease (9).

Because of the perceived usefulness of measuring the flows at various lung volumes during the forced expiratory spirogram, Bartlett and Phillips adopted a flow vs. volume rather than volume vs. time format for this test (12). The advantages of this format include ability to (a) read instantaneous flows at any lung volume; (b) obtain all previous spirographic parameters (a timing mark can be provided at 1-sec intervals for calculating timed volumes); (c) qualitatively interpret the shape of the expiratory flow-volume loop more easily than in a spirogram. For instance, a flow-volume loop with a long, upwardly concave "tail" indicates that a large portion of FVC at low lung volume is exhaled at low flow rates, heralding small airways obstruction (Fig. 2).

The instantaneous flows read from the expiratory flow-volume loop usually include $FEF_{50\%}$, $FEF_{75\%}$, and $FEF_{max}$. $FEF_{50\%}$ and $FEF_{75\%}$ are flows computed after 50% and 75%, respectively, of VC have been exhaled, and therefore qualify as being measured during the "effort-independent" period. As such, they are considered tests of small airways dysfunction, although most studies have found that they exhibit greater inherent variability than "averaging" parameters like $FEF_{25-75\%}$ (13, 14). Such variability requires a greater decrease from the normal predicted value before a given test can be called statistically abnormal, and reduces the sensitivity of the test.

The measurement of $FEF_{max}$, also known as peak flow, is of particular interest in the clinical management of asthma. It is defined as the largest flow rate recorded during the first 15% of exhaled volume, and must be sustained for at least 10 milliseconds. $FEF_{max}$ declines when a major degree of airways obstruction is present, and correlates well with $FEV_1$. Conversely, it is generally normal when asthma is in remission. $FEF_{max}$ has been found useful in two clinical situations: predicting the onset of an asthma attack at home, and objective monitoring of asthma severity in the physician's office, emergency room, or hospital. In all of these situations $FEF_{max}$ is usually measured without formal spirometry, using inexpensive, portable apparatus of the sort first proposed by Wright (15). A variety of these devices are now available, possessing varying degrees of measurement accuracy (16).

$FEF_{max}$ is important as an objective assessment of asthma severity, principally because both subjective assessment by the patient and even objective assessment (physical findings) by the physician have been found wanting. McFadden and associates (17) found that most patients recovering from an acute asthmatic attack became asymptomatic even when $FEF_{max}$ was still markedly abnormal. Physical findings were also frequently normal while a high degree of obstruction was still present, supporting Kelsen's data (18) that patients are often discharged from emergency room care with significant obstructive impairment remaining, contributing to early relapse. Thus, monitoring of $FEF_{max}$ can be a simple, effective method for assessing severity and clinical progress of a patient with asthma in the medical facility.

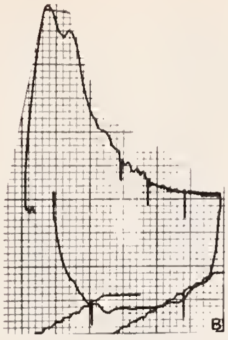


Fig. 2. Maximal flow-volume loop (expiration and inspiration) depicting small airways obstruction. Vertical axis represents flow (expiration is upward), horizontal axis volume (TLC to the left and RV at the right). Note that FVC and all instantaneous flows can be read from curve. Vertical, downward pulses on curve are timing marks representing 1, 2, and 3 seconds, allowing calculation of timed volumes (eg, $FEV_{1.0}$). This loop exhibits concave-upward "tail" resulting from reduced flows at low lung volume typical of small airways obstruction. *Reproduced* with permission from Miller A, ed. Pulmonary function tests in clinical and occupational lung disease. Orlando: Grune and Stratton, 1986:63.

The monitoring of $FEF_{max}$ at home has also become accepted as helpful, principally in the care of the "brittle" asthmatic. The patient is instructed to measure $FEF_{max}$ on a regular schedule, usually once or twice daily. If this measured value falls below some trigger value, the patient should seek medical advice, so that a change in medical regimen can then be recommended to prevent progression to an acute episode. The trigger value can be selected as a percentage of the predicted value for each patient, or on the basis of historical data for a given patient. Taplin and Creer (19) first proposed the latter method based on conditional probability, and it has subsequently been refined and tested by Harm and colleagues (20). In the latter study, data were collected for 60 days, consisting of twice-daily $FEF_{max}$ and time of occurrence of an asthmatic attack. A variety of possible $FEF_{max}$ values were selected, and two conditional probabilities calculated for each: (a) probability of an asthmatic attack occurring in a 12-hour period after an $FEF_{max} \leq$ the selected value; and (b) probability of an attack occurring in a 12-hour period after an $FEF_{max} >$ the selected value. The ratio of these two probabilities ("hits to misses ratio") for each value of $FEF_{max}$ indicates the error rate inherent in choosing an $FEF_{max}$ trigger point at that value. A value of $FEF_{max}$ was then chosen which maximized this ratio yet provided the best prediction of an asthma attack.

Home monitoring of $FEF_{max}$ has also been suggested as a method to improve the correlation between patients' symptoms and objective pulmonary function. In this hypothesis, patients measuring $FEF_{max}$ at home would learn to identify certain symptoms associated with declining function, and appropriate intervention could then be sought. Current published data are contradictory

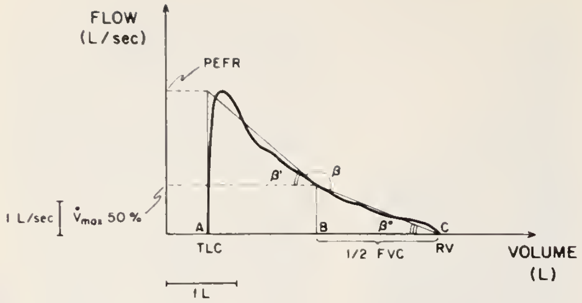


Fig. 3. Definition of angle β on the maximal expiratory flow-volume loop. *Reproduced* with permission from Kapp MC, Schachter EN, Beck GJ, Maunder LR, Witek TJ Jr (ref. 26), 800.

about whether this training effect takes place (21, 22).

## Novel Methods for Analyzing the Forced Expiratory Spirogram

The simple parameters described above by no means encompass all of the possible methods by which the forced expiratory spirogram can be analyzed. Novel techniques for detecting obstructive impairment have been especially popular as subjects of research. These methods may be divided into three general categories:

- analysis of flow-volume loop shape
- time or frequency-domain analysis
- spirometric evidence of air trapping

**Analysis of Flow-Volume Loop Shape.** Green and colleagues developed the first index of expiratory flow-volume loop shape, which described the extent the terminal portion of the curve deviated from a straight line (23). They reported that this index increased with age, but did not provide details of the computation. Mead (24) developed an analytic technique which he called the slope ratio (SR) method. For each volume following peak flow, the ratio between instantaneous slope (a tangent to the line at that point) and average slope from that volume to the origin (a chord drawn between the point on the flow volume loop and the origin) is computed. Mathematically, this is equivalent to:

$$SR(V) = (d\,\dot{V}/V) \,/\, (\dot{V}/V) \qquad (3)$$

If the curve at a given volume is parallel to the straight line drawn to the origin from that point, SR will be unity. If the curve is steeper than the straight line (convex upward) SR will be less than one; if the curve is shallower (concave upward—the "tail" indicative of airways obstruction), SR will be greater than 1. Mead was able to analyze the SR vs. volume curves thus obtained in terms of homogeneous and nonhomogeneous models of lung emptying, and including airway closure. However, the method proved to be an insensitive indicator of obstruction in bronchial asthma (specifically, exercise-induced bronchospasm).

An additional problem associated with the SR method involves how to use the data thus obtained. The result is another plot, this time of SR vs. volume, rather than a single parameter which can be more easily used in comparisons. Two other methods are available which result in single parameters describing the degree of upward concavity of the flow-volume curve. One index, 1/2 ($FEF_{50\%}/FEF_{75\%}$), was arrived at by calculating the ratio between chord slopes drawn from points after 50% and 75% of VC were exhaled (25):

$$\begin{aligned}(FEF_{50\%}/0.5\ VC)/(FEF_{75\%}/0.25\ VC) \\ = 1/2\,(FEF_{50\%}/FEF_{75\%})\end{aligned} \qquad (4)$$

This ratio, which is unity for a straight curve and increases with concavity, has attracted little notice in the literature.

The most recently described parameter, angle β (26), has seen slightly more use (Fig. 3). Graphically, the angle is calculated by drawing lines between the $FEF_{50\%}$ point on the flow-volume curve and both a point at TLC corresponding to the height of $FEF_{max}$ and the point on the volume axis at RV. Angle β can also be calculated directly from various flow-volume parameters:

$$\begin{aligned}\beta(\text{degrees}) = 180° - \tan^{-1}[(FEF_{max} \\ - FEF_{50\%})/(½\ FVC)] \\ + \tan^{-1}(FEF_{50\%}\,/\,½\ FVC)\end{aligned} \qquad (5)$$

As is obvious from inspection of the graphical analysis, β is 180° when the descending limb of the curve inscribes a straight line; greater than 180° with convexity upward; and less than 180° with the upward concavity typical of obstructive disease. Kapp and colleagues (26) investigated angle β in a cross-sectional study of 5,140 individuals in 3 communities. They found that β correlated significantly with pack-years of smoking, falling as tobacco exposure increased. In subjects having a history of asthma or wheezing β was significantly lower; β also correlated strongly with other standard indices of obstruction such as $FEV_1$, but had less between-subject variability than any obstructive parameter studied. A subsequent study by the same investigators (27) utilized β in an analysis of pulmonary function in 477 cotton textile workers. They again showed a tendency for $\beta$ to decline in bronchial asthma. Interestingly, there was no decline in $\beta$ when cotton textile workers were compared to control subjects. The authors hypothesized that $\beta$ may be uniquely sensitive to small airways obstruction, while the site of damage in byssinosis may reside elsewhere.

**Time-Domain Analysis.** So-called time domain analysis of the spirogram was first introduced by Fish and colleagues in 1974 (28) and extended to frequency domain analysis by Permutt and Menkes in 1979 (29). Time-domain analysis relies on the concept of *transit time,* or the time required for each small parcel of vital capacity ($\Delta V$) to arrive at the mouth and be measured during the forced exhalation. Fig. 4 depicts a conventional spirogram, except that normalized volume (V/VC) is used for the ordinate. In time-domain analysis, such a graph represents a cumulative distribution function, in which the fraction of vital capacity arriving with a transit time $\leq t$ is depicted. A large body of mathematics is available for the analysis of such cumulative distribution functions; one of the most powerful relies on the computation of moments, in which the distribution function is integrated in powers of *t:*

$$\alpha_r = \int_0^{VC} \frac{t^r dV}{VC} \qquad (6)$$

where $\alpha_r$ is the r$^{th}$ moment of time about the origin.

Certain statistical descriptors of the shape of the cumulative distribution function are available using the calculated moments. For instance, $\alpha_1$ is the same as the mean transit time (the average value for all of the transit times of the distribution, also called "effective time" in some literature) and is equivalent to the time constant for lung emptying if the spirogram could be described by a single exponential function. The standard deviation of the distribution of transit times is represented as $(\alpha_2)^{1/2}$ (moments are conventionally expressed as their $r_{th}$ root). The ratio $(\alpha_2)^{1/2}/\alpha_1$ yields a specific value ($\sqrt{2}$) if the distribution function is a single exponential, and will exceed this value as the function becomes more disperse. This is equivalent to the function being made up of a sum of exponential functions with different time constants.

Permutt and Menkes suggested several advantages of analysis using moments (29). First, moments (especially higher order moments) stress events during the terminal portion of the spirogram, and thus should be more sensitive indicators of small airways obstruction. Second, moments should exhibit greater precision than parameters measured at single points in time, since they are calculated from the entire spirogram. Third, it should be possible to characterize the shape of the entire spirogram using only the first 4 moments. Another possible advantage resides in the fact that the moments are calculated from the normalized spirogram, and should be insensitive to concomitant restrictive impairment.

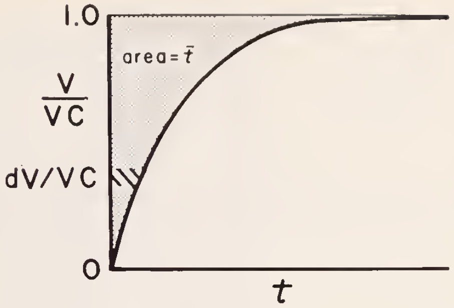


Fig. 4. Forced expiratory spirogram modified to represent cumulative distribution function. Fraction of VC (actually, FVC) is depicted on ordinate, time on abscissa. Each point on curve represents fraction of VC delivered with transit time of $\leq$t. Shaded area, $\tau$, represents mean transit time, calculated by integrating dV/VC against *t.* *Reproduced* with permission from Permutt S, Menkes HA (ref. 29), 117.

Not surprisingly, most investigations using moment analysis have involved attempts to detect early obstructive disease in smokers. A few studies, however, have involved detection of disease in asthmatics. Fish and colleagues (28) measured $\alpha_1$ and $(\alpha_2)^{1/2}/\alpha_1$ in 6 asthmatics before and after antigen challenge, and 3 normal subjects before and after methacholine challenge. In all subjects $\alpha_1$ increased following induction of bronchospasm, consistent with the slowing of expiratory airflow from obstructive impairment. However, $(\alpha_2)^{1/2}/\alpha_1$ did not increase, which could be interpreted as evidence for a uniform effect of the challenge substance on airway caliber (homogeneity of lung units rather than heterogeneity). Similarly, Liang and coworkers (30) studied various spirometric indices, including $\alpha_1$ and coefficient of variation (an index of dispersion similar to $(\alpha_2)^{1/2}/\alpha_1$) in 51 asthmatic children in clinical remission. They found $\alpha_1$ to be the most sensitive indicator of abnormality in this group; coefficient of variation was generally normal, although the group mean value was slightly elevated compared to their reference value.

**Frequency-Domain Analysis.** Findings such as these suggest a relatively uniform distribution of pathologic changes in asthmatic airways. Permutt and Menkes, however, have suggested that

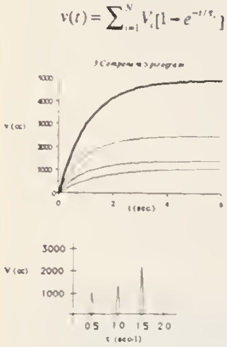
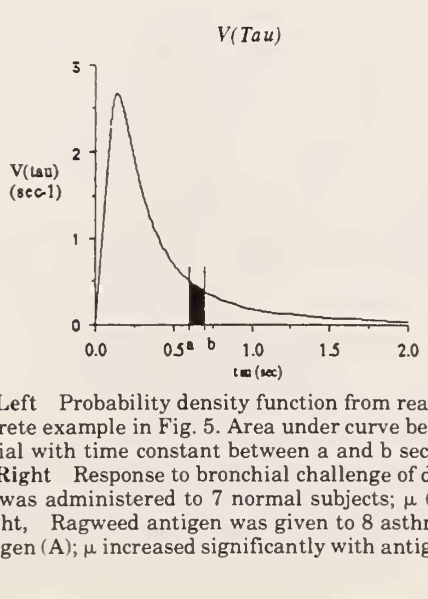


**Fig. 5.** Separation of forced expiratory spirogram into sum of exponential functions. Middle panel, exponential functions with 3 different time constants (thin lines) and spirogram made up of sum of these exponentials (thick line). Top panel, mathematical representation of this process; bottom panel, probability density function for this spirogram. *Courtesy of* M. Lacker.

an index of dispersion such as $(\alpha_2)^{1/2}/\alpha_1$ can fall as $\alpha_1$ rises as a consequence of truncation error (failure to analyze the full expiration) during the spirometric maneuver, which they believe is a common occurrence (29). In order to minimize this type of error, they proposed an extension of moment (time-domain) analysis to what may be better termed frequency-domain analysis.

Mathematically, each transit time defines an exponential function with a given time constant, and the sum of these exponential functions makes up the entire (normalized) spirogram. Fig. 5 depicts a spirogram made up of 3 exponential functions: the function with $\tau$ (time constant) = 0.5 contributes ~1.0 L to the total VC; $\tau$ = 1.0 contributes ~1.5 L; and $\tau$ = 1.5 contributes just over 2 L. One can then construct a so-called probability density function (bottom panel, Fig. 5) which defines the likelihood that expired air will be attributable to an exponential function with a given $\tau$. Anatomically, each $\tau$ reflects the myriad of lung units possessing the same mechanical time constant ($R_{aw} \times C_l$). Thus, the probability density function (also known as a frequency distribution, hence frequency domain-analysis) for a typical spirogram will take the form of a continuous function (Fig. 6). It is suggested that the attributes of this function (mean, standard deviation, and so on) will describe the shape of the spirogram and mechanical aspects of the system which produced it. For instance, a frequency distribution with a high mean value implies that much of the VC results from exponential functions with long time constants, an indicator of obstructive impairment. A frequency distribution with a large standard deviation indicates that VC emanates from exponentials with greatly varying time constants, indicating heterogeneous airway disease.

The manner in which this type of analysis sidesteps the truncation-error issue involves the method by which the frequency distribution is calculated. Both methods currently available result in the fitting of a new equation to the spirometric data points, such that the new equation follows the shape of the spirogram but is constrained to reach the abscissa asymptotically as t increases. Permutt and Menkes (29) used an iter-



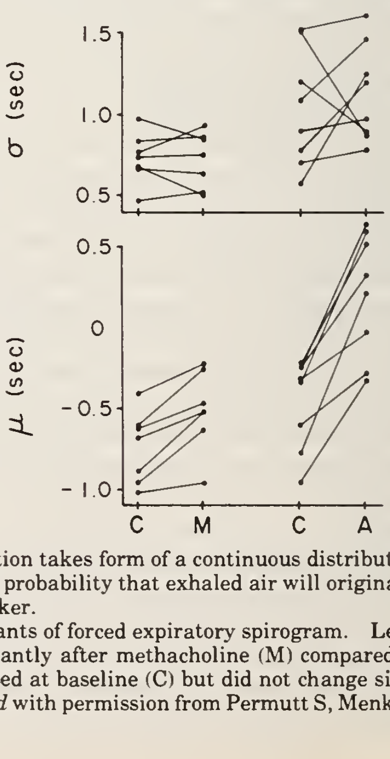


**Fig. 6.** **Left** Probability density function from real spirogram. Such a function takes form of a continuous distribution, rather than discrete example in Fig. 5. Area under curve between a and b represents probability that exhaled air will originate from an exponential with time constant between a and b seconds. *Courtesy of* M. Lacker.

**Fig. 7.** **Right** Response to bronchial challenge of distribution of time constants of forced expiratory spirogram. **Left,** methacholine was administered to 7 normal subjects; μ (mean) increased significantly after methacholine (M) compared to control (C). **Right,** Ragweed antigen was given to 8 asthmatics; σ (SD) was elevated at baseline (C) but did not change significantly after antigen (A); μ increased significantly with antigen challenge. *Reproduced* with permission from Permutt S, Menkes HA (ref. 29), 139.

ative method assuming a lognormal form for the frequency distribution; we have published a method not requiring an assumption as to the form of the distribution function, which uses the inverse Laplace transform of a function fitted to the spirogram (31).

Frequency-domain analyses have mostly been applied to detecting early airways disease attributable to smoking. However, Permutt and Menkes did determine mean and standard deviation in a group of 7 normal subjects before and after methacholine inhalation, and 8 asthmatics before and after ragweed antigen challenge (29). Both showed a striking elevation in mean following exposure, consistent with airways obstruction; neither exhibited an increase in standard deviation, confirming their earlier observation in a similar group that airways changes were relatively homogenous (Fig. 7). Interestingly, the standard deviation was somewhat elevated at baseline in the asthmatics, suggesting heterogeneous pathologic abnormalities in this disease which were not exaggerated by acute bronchospasm.

**Spirometric Evidence of Air Trapping.** Finally, some investigators have been able to demonstrate air trapping using the forced expiratory spirogram. Dynamic air trapping occurs in obstructive lung disease when the pressures exerted on the diseased airway during forced expiration are such that premature closure of these airways takes place, "trapping" the alveolar air distal to the closure. Such closure may also occur when elastic recoil of the lung parenchyma is reduced, such as in pulmonary emphysema, since this recoil force helps to maintain airway patency. In the past, dynamic air trapping was demonstrated spirometrically by performing either of two sets of maneuvers:

**Slow vital capacity (SVC) and FVC.** The difference between SVC and FVC is a measure of dynamic air trapping, and when a marked disparity occurs, pathologic air trapping may be presumed.

**FVC by spirometry and body plethysmography.** The subject performs a standard forced expiratory spirogram, with volume change determined both spirometrically (measures exhaled volume) and using the body plethysmograph (measures the reduction in lung volume). A disparity between these volumes (spirometric FVC being less than plethysmographic FVC) indicates compression of trapped intrathoracic air, as was successfully demonstrated by Charan and colleagues (32) in both asthmatics and smokers.

Dynamic air trapping may be demonstrated more simply in some asthmatics. Miller (33) has described changes in "forced expiratory reserve volume" attributable to this effect. Expiratory reserve volume (ERV) is conventionally defined as the difference in volume between functional residual capacity (FRC; the end of a normal tidal breath) and residual volume, as measured during a slow expiration to RV (slow vital capacity maneuver). ERV can just as well be measured from the forced expiratory spirogram, as long as several tidal breaths are recorded previously to define the FRC point; Miller termed this the forced ERV (FERV) and found it to be reduced or even negative in 25 asthmatic subjects (Fig. 8). Interestingly, all the patients were selected to have normal $FEV_1/FVC$ ratios; only 13 had prolonged $FET_{25-75\%}$.

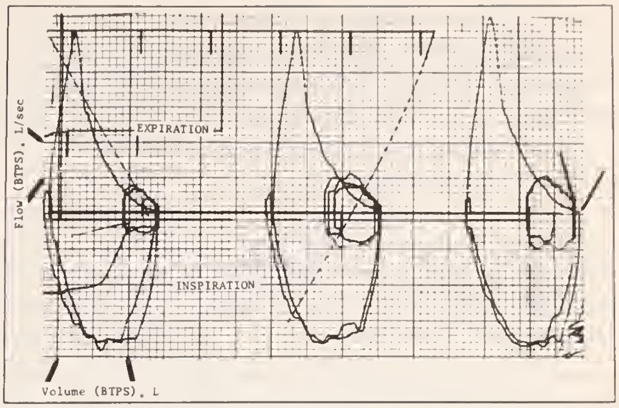


Fig. 8. Three maximal inspiratory/expiratory flow volume loops in patient with bronchial asthma, demonstrating reduced forced ERV sign. Flow is on vertical axis, volume on horizontal; expiratory flow is upward. Expiration from TLC to RV proceeds from left to right. Small loops within each effort are tidal breaths, which define FRC as farthest excursion of each loop to right. Maximal expiratory curve on right extends only 80 mL below FRC (FERV = 80 mL). Other maximal expiratory curves terminate at FRC (FERV = 0 mL). *Reproduced* with permission from Miller A (ref. 33), 89.

Dynamic air trapping may also be demonstrated spirometrically using a maneuver other than on forced expiration. Maximum voluntary ventilation (MVV; also known as maximum breathing capacity, MBC) is defined as the highest minute ventilation that can be sustained over a 10–15–sec period. The total volume either inhaled or exhaled through a spirometer for a 10–15–sec period is measured by exhorting the subject to breathe in and out "as hard and fast as you can." MVV is then calculated by expressing this volume as if it had been sustained for a full minute. MVV is thus really a flow rate (volume/time) and is reduced in obstructive disorders such as bronchial asthma; it correlates well, in fact, with $FEV_1$.

When the MVV maneuver is displayed as the full volume vs. time tracing, evidence of dynamic air trapping consists of a rise in FRC during the course of the test. This occurs when inhaled air is trapped during exhalation and does not appear at the mouth; it is similar to the "autopeep" effect commonly observed in obstructed patients on mechanical ventilation.

### Pharmacologic Manipulations Involving Forced Expiratory Spirogram

The simplest manipulation involves forced expiratory spirometry before and after administration of a bronchodilator aerosol. A significant improvement in spirometric parameters following bronchodilator medication has been considered diagnostic of bronchial asthma (34), although not all asthmatics will respond rapidly enough to achieve this. Commonly, a beta$_2$-specific adrenergic drug is utilized, administered using a standard metered-dose inhaler or compressed-air nebulizer.

The inherent variability of spirometric parameters must be considered in interpreting whether a post-bronchodilator change is significant. Pennock and coworkers (35) summarized data appearing in the literature, as well as their own, to arrive at coefficients of variation (CV) for spirometric tests which have become generally accepted. In normal subjects, CV (within-day variation) approximates 3% for FVC and FEV$_1$, and 8% for FEF$_{25-75\%}$. In patients with obstructive lung disease, CV increases to 6.7% for FVC, 8.1% for FEV$_1$, and 14% for FEF$_{25-75\%}$. Translating these figures into 95% confidence levels for a significant within-day change, FVC must increase by 11%, FEV$_1$ by 13%, and FEF$_{25-75\%}$ by 23% in obstructive disease. A subsequent study by Sourk and Nugent (36) tested this more directly by administering inhaled placebo to patients referred for pulmonary function testing. The upper 95% confidence limits obtained were 14.9% for FVC, 12.3% for FEV$_1$, and 45.1% for FEF$_{25-75\%}$. These studies thus closely approximate the commonly used guidelines requiring a 15% increase in FVC or FEV$_1$ to confirm a positive bronchodilator response. One should use these guidelines with care, however, in patients with small baseline values for FVC or FEV$_1$. Miller (37) points out that a 15% increase in FEV$_1$ can, at times, represent an absolute increase of only 100 cc or so and recommends a minimum absolute increase for FVC or FEV$_1$ of 200 cc, in addition to the 15% criterion.

Asthmatics may respond to bronchodilator medication by increasing either FEV$_1$, FVC, or both. Intuitively, one might expect only a flow response, or an increase in FEV$_1$. Ramsdell and Tisi (38) demonstrated an increase in FVC ("volume-response") in 46 of 241 patients undergoing bronchodilator testing. These volume-responders generally exhibited more airways obstruction and hyperinflation at baseline than the flow-responders, leading the authors to conclude that reduced lung elastic recoil, and consequent dynamic airway closure, explained their findings. Paré and coworkers (39) obtained similar results in a group of 15 adult asthmatics, but also analyzed for site of obstruction by performing helium-oxygen flow-volume loops (see below). There was increased evidence for small airways obstruction in the volume-responders following bronchodilator medication, suggesting that this treatment opened a population of peripheral airways that were completely occluded at baseline. Lung distal to these occluded airways was then available to contribute volume to the measured FVC, but remained narrow enough to locate the majority of obstruction peripherally.

A consequence of this volume-response to bronchodilators is the difficulty sometimes encountered when mean or instantaneous flows are used to gauge bronchodilator responsiveness. If these flows are referenced to the post-bronchodilator lung volume, they will frequently be measured at a lower absolute lung volume compared to the same measurement done prior to medication. Since flows at lower lung volumes in asthmatics are generally reduced, this will tend to obscure any actual improvement in airway caliber. A more accurate method involves measuring these flows at isovolume to the baseline curve, either using the body plethysmograph to obtain absolute volume, or by assuming TLC to be constant and determining the flows at an equal volume below TLC for both baseline and post-bronchodilator curves (40). Light and coworkers (41), however, suggest that FEV$_1$ is still the most sensitive index of bronchodilation, and performing the isovolume flow determination may not be worth the added difficulty.

A more complex pharmacologic manipulation involves the administration of provocational agents aimed at inducing bronchospasm in susceptible individuals. The cholinergic agent methacholine chloride (mecholyl) is most often employed for general diagnostic use; histamine can yield similar results but is sometimes less well tolerated (42). Distilled water aerosol (43), cold air (44), and exercise (45) have also been used in this regard, but are less well standardized. Specific antigens can be utilized to determine the etiology of asthmatic episodes related to workplace or other environmental exposure (46). Leukotriene D, a component of slow reacting substance of anaphylaxis, has also been used in the research setting (47).

When an aerosolized agent such as methacholine is used, it may be administered using a breath-activated dosimeter or by tidal breathing from a facemask. The dosimeter delivers a precise quantity of aerosol during each inhalation (48), and is probably more desirable for research studies. Tidal breathing from a facemask requires less complex apparatus, has been shown to offer similar reproducibility, and is quite satisfactory for use in the clinical laboratory (49). Pulmonary function (usually spirometry and sometimes specific airways conductance, $sG_{aw}$) is measured at baseline and after inhalation of the diluent (phosphate-buffered saline) alone; increasing concentrations of the provocational agent are then administered with pulmonary function studies repeated after each dosage level. Conventionally, 5 dosimeter inhalations or 2 minutes of tidal breathing from the facemask constitute a given dosage level; methacholine concentrations generally start at 0.075 mg/mL and are doubled each time (48, 49). Abbreviated dosage schedules have been described and may be utilized in certain clinical situations (50). In the standard protocol, $FEV_1$ is plotted against drug dose, expressed as cumulative units. When using methacholine, one inhalation of a 1 mg/mL concentration is defined as one unit. Thus, 0.375 units are delivered after 5 inhalations of the first dosage level of 0.075 mg/mL, 1.125 units after the second 5 inhalations of 0.15 mg/mL, etc. This is sometimes simplified such that only the concentration of drug at each level is recorded.

Just as in the standardization of the bronchodilator response, a magnitude of change in $FEV_1$ must be defined to represent a significant result. Unlike bronchodilator testing, a 20% reduction in $FEV_1$ is considered as significant, even though this is somewhat greater than the 95% confidence limit for the test, as detailed above. The control value is taken as the $FEV_1$ following the 5 breaths of diluent, and the dose (in cumulative breath units) at which a 20% reduction occurs is found graphically. This value is termed the provocational dose for a 20% reduction in $FEV_1$, or $PD_{20}$ for $FEV_1$. If only the concentration at each level is recorded, then the provocational concentration for a 20% reduction ($PC_{20}$) is obtained. Other pulmonary function parameters ($sG_{aw}$, FVC, $FEF_{max}$, $FEF_{25-75\%}$) may also be used, generally with different significance limits, although $FEV_1$ is the most standardized.

The $PD_{20}$ or $PC_{20}$ value thus obtained can be used clinically to identify patients with hyperreactive airways. Malo and coworkers have published reference values for $PC_6$ and $PC_{20}$, suggesting a $PC_{20} > 8$ mg/mL as a reasonable level for separating normal from hyperreactive subjects (51). Such a finding can then be used to guide treatment. For instance, Corrao and colleagues described a group of patients complaining of chronic cough, who were found to be hyperreactive on methacholine challenge (50). Treatment with bronchodilators provided substantial symptomatic relief, suggesting these patients were subclinical asthmatics.

Whether airways hyperreactivity can be equated to a diagnosis of bronchial asthma is somewhat more controversial. Studies of bronchial hyperreactivity compared with clinical symptoms and signs suggest that challenge testing will identify only a subgroup of patients with asthma; conversely, some subjects with bronchial hyperresponsiveness cannot be identified clinically as asthmatic (52). However, "all that wheezes is not asthma," and the diversity of disease states that can cause asthmalike symptoms and signs is considerable. Kelso, for instance, described a group of patients with asthmatic symptoms who responded to methacholine challenge testing with spirometric evidence of extrathoracic, upper airway obstruction (reduced inspiratory flows) (53). It is likely, therefore, that the clinical diagnosis of bronchial asthma describes a heterogeneous group of disorders; those with airways hyperreactivity comprise only one element of this population.

Helium-oxygen (He-$O_2$) flow-volume curves may be considered in a discussion of pharmacologic manipulations of the forced expiratory spirogram. Although the breathing of He-$O_2$ is more frequently done for diagnostic purposes, as will be described, some have advocated it as a therapeutic modality as well (54).

Schilder and coworkers (55) originally described the effect of breathing gas mixtures of different densities and viscosities on the expiratory flow-volume curve. They noted that an effort performed after replacing thoracic gas with a helium-oxygen mixture (less dense than air) resulted in a shift of the initial portion of the flow-volume curve upwards, representing higher flow

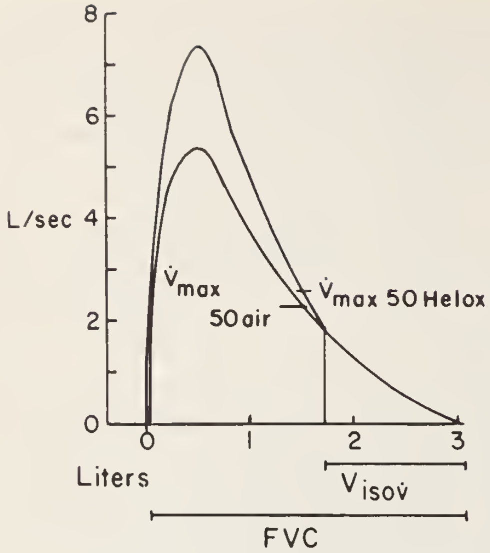


**Fig. 9.** Maximal expiratory flow-volume curves breathing air (lower curve) and helium-oxygen (upper curve). $\dot{V}_{max}$ 50% points for air and helium-oxygen are identified, as is volume of isoflow ($V_{iso}V$) where curves superimpose. *Reproduced* with permission from Ashutosh K, Glenn Mead PT, Dickey JC Jr, Berman P, Kuppinger M. Density dependence of expiratory flow and bronchodilator response in asthma. Chest 1980; 77:70.

rates (Fig. 9). This initial portion of the curve is governed by turbulent flow through the larger airways (see discussion of the equal pressure point, above), and the increase in flow rate observed was consistent with fluid mechanical theory which has shown that resistance to turbulent airflow is proportional to gas density. Despas and colleagues subsequently used these findings to demonstrate differing sites of airflow resistance in patients with asthma and chronic bronchitis (56). They looked at the $FEF_{50\%}$ (usually expressed as $\dot{V}_{max}$ 50% in this context) for air and helium oxygen curves in each subject, and found that these values did not differ in chronic bronchitics and a few asthmatics, indicating that the majority of airflow resistance resided in the small airways in these individuals. In most asthmatics, $\dot{V}_{max}$ 50% increased when breathing helium-oxygen, localizing the site of flow resistance to the larger airways. Dosman et al. (57) then refined these measurements to arrive at the two indices currently in use: $\Delta\dot{V}_{max}$ 50%, defined as

$$\Delta\dot{V}_{max}\,50\% = \frac{\dot{V}_{max}\,50\%\,(\mathrm{He}) - \dot{V}_{max}\,50\%\,(\mathrm{Air})}{\dot{V}_{max}\,50\%\,(\mathrm{Air})} \times 100 \qquad (7)$$

and $V_{iso}\,\dot{V}$ (volume of isoflow), which is defined as the volume (expressed as %VC from RV), at which the helium-oxygen and air flow-volume curves join (Fig. 9). With greater density-dependence of airflow, $\Delta\dot{V}_{max}$ 50% rises, and $V_{iso}\,\dot{V}$ declines; Dosman was thus able to demonstrate a significant shift of the site of airflow obstruction to the small airways in smokers.

Tests of density dependence of maximal expiratory airflow have been most useful in the research laboratory. One can determine the site of bronchodilatation when various pharmacologic agents are given: atropine, for instance, may preferentially affect the larger airways while isoproterenol acts on the small airways (58). Similarly, the site of pharmacologic bronchoconstriction can be determined: Pichurko et al. demonstrated a prominent large airway response to leukotriene $C_4$ compared to the constriction of small airways induced by histamine (59).

Others workers have used these tests to define subtypes of bronchial asthma. Eliasson and Zuwalack, for example, showed that, in asthmatics requiring hospitalization for their illness, those with predominantly large airways obstruction improved more rapidly than those with small airways disease (60). As might be expected from the heterogeneity of the site of airflow resistance in asthma (56), these tests are probably not useful in studies attempting to identify early or clinically inapparent disease (61).

## Nonspirometric Tests of Pulmonary Function

These tests have generally been less clinically useful and require more complicated apparatus than spirometric tests, and are summarized here only briefly.

**Pulmonary Mechanics.** Airway resistance ($R_{aw}$) may be conveniently measured in the body plethysmograph (62). It is elevated during acute episodes of bronchial asthma, and frequently lags behind the improvement in symptoms induced by treatment (17). $R_{aw}$ may underestimate the degree of airways obstruction: $R_{aw}$ declines with increasing lung volume, such as the hyperinflation associated with acute exacerbations of asthma. Specific airways conductance ($sG_{aw}$) corrects for this by normalizing to the lung volume at which the measurement is made:

$$sG_{aw} = (1/R_{aw})/\mathrm{FRC} \qquad (8)$$

For this reason, $sG_{aw}$ (rather than $R_{aw}$) has been used to monitor bronchial challenge testing (see above).

Total respiratory system resistance ($R_{rs}$) can also be used to assess obstructive impairment in asthma, since an increase in this value generally represents a change in the $R_{aw}$ component. Apparatus is available for measuring $R_{rs}$ using the forced oscillation method, including a system for performing automated methacholine challenge testing (63, 64).

Volume history has been shown to affect airways resistance in asthmatics but not normal subjects. A deep inspiration preceding measurement of $R_{aw}$ in asthmatics will cause a transient increase in this value (65). The mechanism was thought to be vagally mediated but this has recently been disputed (66).

Static lung compliance—$C_{st}(l)$—remains normal or is slightly reduced during exacerbations of asthma (67), while the lung pressure-volume curve shifts upward and to the left, resulting in lower transpulmonary pressures for a given degree of lung inflation (67, 68). The mechanism behind this has never been satisfactorily explained, although some authors implicate the phenomenon of tissue stress relaxation (67). Whatever the cause, the alteration in pressure-volume relationship probably contributes to the hyperinflation associated with asthmatic episodes (see below).

Dynamic compliance—$C_{dyn}(l)$—becomes frequency dependent (its value declines with increasing respiratory rate) in disorders associated with small airways obstruction. A landmark theoretical analysis by Otis et al. implicated the presence of a heterogeneous population of lung units with different time constants as the basis for this phenomenon (69). They demonstrated the development of frequency-dependence of compliance with histamine-induced bronchospasm in normal individuals, and this test has subsequently become the "gold standard" against which other tests for small airways dysfunction are judged (70).

Other tests of distribution of ventilation (single or multiple breath nitrogen washout, helium wash-in) become abnormal with exacerbations of asthma, and may remain abnormal even into remission (71). The theoretical basis behind this result is identical to that causing frequency-dependence of compliance (72).

**Static Lung Volumes.** Static volumes (TLC, FRC, RV) are those volumes and capacities which cannot be measured by spirometry alone because they include residual volume as one of their components. They can be measured by a gas dilution method (single or multiple breath nitrogen washout or helium wash-in) or in the body plethysmograph. The gas dilution methods, especially the single breath techniques, may seriously underestimate lung volume due to failure of the gas to equilibrate with poorly-ventilated lung units (73). Body plethysmography may overestimate lung volume; two mechanisms have been suggested: inhomogeneity of pressure changes due to the distribution of time constants in the obstructed individual (74), or the effect of the compliance of the extrathoracic airway (75, 76).

TLC, FRC, and RV all increase during exacerbations of asthma, and may remain elevated even during remission (17). Mechanisms for this include changes in lung elastic recoil as noted above; air trapping; and increased chest wall elastic recoil. The latter appears to be due to tonic *inspiratory* muscle activity persisting during *expiration*, which has been demonstrated during acute bronchospasm (77, 78).

**Gas Exchange.** The ventilation-perfusion mismatching caused by obstruction in the small airways results in hypoxemia; interestingly, beta-adrenergic medication may worsen hypoxemia by increasing blood flow to poorly ventilated lung units (79). Except in severe status asthmaticus with respiratory muscle fatigue, $P_aCO_2$ declines with the onset of bronchospasm (80). Tests of respiratory center output are elevated (81), which is partly a consequence of hypoxemia but may be more related to stimulation of vagally mediated irritant or stretch receptors (82).

Despite abnormal ventilation-perfusion relationships, single breath carbon monoxide diffusing capacity ($D_LCO_{SB}$) is normal or even increased (83, 84), in contrast to the severely reduced values found in pulmonary emphysema.

## Summary

Despite the myriad of tests available to characterize bronchial asthma, the simple forced expiratory spirogram remains the most straightforward and reliable technique to diagnose the disorder and gauge its severity. When combined with bronchodilator and bronchial challenge testing, a wide range of clinical situations can be evaluated, even in the office setting. A well-performed forced expiratory spirogram remains the cornerstone of pulmonary function testing; the emergence of the additional analytic techniques described holds the promise of even greater utility in the future.

### References

1. American Thoracic Society. Standards for the diagnosis and care of patients with chronic obstructive pulmonary disease (COPD) and asthma. Am Rev Respir Dis 1987; 136:225–243.
2. Murlas CG. Pathogenesis of airway hyperreactivity. Chest 1988; 93:1278–1280.

3. Palmer KNV, Kelman GR. Pulmonary function in asthmatic patients in remission. Br Med J 1975; 1:485–486.
4. Dunnill MS. The pathology of asthma, with special reference to changes in the bronchial mucosa. J Clin Pathol 1960; 13:27–33.
5. Tiffeneau R, Pinelli A. Air circulant et air captif dans l'exploration de la fonction ventilatrice pulmonaire. Paris Med 1947; 37:624.
6. Gaensler EA. Analysis of the ventilatory defect by timed vital capacity measurements. Am Rev Tuberc 1951; 64:256–278.
7. Morris JF, Temple WP, Koski A. Normal values for the ratio of one-second forced expiratory volume to forced vital capacity. Am Rev Respir Dis 1973; 108:1000–1003.
8. Fry DL, Hyatt RE. Pulmonary mechanics: a unified analysis of the relationship between pressure, volume and gasflow in the lungs of normal and diseased human subjects. Am J Med 1960; 29:672–689.
9. Leuallen EC, Fowler WS. Maximal midexpiratory flow. Am Rev Tuberc 1955; 72:783–800.
10. Berend N, Woolcock AJ, Marlin GE. Correlation between the function and structure of the lung in smokers. Am Rev Respir Dis 1979; 119:695—705.
11. Wright JL, Lawson LM, Paré PD, Kennedy S, Wiggs B, Hogg JC. The detection of small airways disease. Am Rev Respir Dis 1984; 129:989–994.
12. Bartlett RG Jr, Phillips NE. The velocity volume loop: a composite pulmonary function test. Dis Chest 1962; 42:482–488.
13. Cochrane GM, Prieto F, Clark TJH. Intrasubject variability of maximal expiratory flow volume curve. Thorax 1977; 32:171–176.
14. Miller A, Thornton JC, Warshaw R, Bernstein J, Selikoff IJ, Teirstein AS. Mean and instantaneous expiratory flows, FVC and FEV$_1$: prediction equations from a probability sample of Michigan, a large industrial state. Bull Eur Physiopathol Respir 1986; 22:589–597.
15. Wright BM, McKerrow CB. Maximum forced expiratory flow rate as a measure of ventilatory capacity. Br Med J 1959; 2:1041–1046.
16. Eichenhorn MS, Beauchamp RK, Harper PA, Ward JC. An assessment of three portable peak flow meters. Chest 1982; 82:306–309.
17. McFadden ER Jr, Kiser R, de Groot WJ. Acute bronchial asthma: relations between clinical and physiologic manifestations. N Engl J Med 1973; 288:221–225.
18. Kelsen SG, Kelsen DP, Fleegler BF, Jones RC, Rodman T. Emergency room assessment and treatment of patients with acute asthma: adequacy of the conventional approach. Am J Med 1978; 64:622–628.
19. Taplin PS, Creer TL. A procedure for using peak expiratory flow rate data to increase the predictability of asthma episodes. J Asthma Res 1978; 16:15–19.
20. Harm DL, Kotses H, Creer TL. Improving the ability of peak expiratory flow rates to predict asthma. J Allerg Clin Immunol 1985; 76:688–694.
21. Higgs CMB, Richardson RB, Lea DA, Lewis GTR, Laszlo G. Influence of knowledge of peak flow on self assessment of asthma: studies with a coded peak flow meter. Thorax 1986; 41:671–675.
22. Reeder KP, Dolce JJ, Duke L, Raczynski JM, Bailey WC. Peak flow meters: are they monitoring tools or training devices? J Asthma 1990; 27:219–227.
23. Green M, Mead J, Hoppin F, Wohl ME. Analysis of the forced expiratory maneuver. Chest 1973; 63(Suppl): 33S–35S.
24. Mead J. Analysis of the configuration of maximum expiratory flow-volume curves. J Appl Physiol 1978; 44: 156–165.
25. Hyatt RE, Rodarte JR, Mead J, Wilson TA. Changes in lung mechanics. Flow-volume relations. In: Macklem PT, Permutt S, eds. The lung in the transition between health and disease. New York: Marcel Dekker, 1979:73–112.
26. Kapp MC, Schachter EN, Beck GJ, Maunder LR, Witek TJ Jr. The shape of the maximum expiratory flow volume curve. Chest 1988; 94:799–806.
27. Schachter EN, Kapp MC, Beck GJ, Maunder LR, Witek TJ. Smoking and cotton dust effects in cotton textile workers. Chest 1989; 95:997–1003.
28. Fish J, Menkes H, Rosenthal R, Summer W, Norman P, Permutt S. The effect of acute bronchospasm on the distribution of transit times during forced expiration (abstract). Am Rev Respir Dis 1974; 109:700.
29. Permutt S, Menkes HA. Spirometry: analysis of forced expiration within the time domain. In: Macklem PT, Permutt S, eds. The lung in the transition between health and disease. New York: Marcel Dekker, 1979:113–152.
30. Liang A, Macfie AE, Harris EA, Whitlock RML. Transit-time analysis of the forced expiratory spirogram during clinical remission in juvenile asthma. Thorax 1979; 34:194–199.
31. Brown LK, Lacker HM, Pilipski M, Miller A. Analysis of the forced expiratory spirogram by computation of the distribution of expiratory rate-constants (abstract). Chest 1990; 98:34S.
32. Charan NB, Hildebrandt J, Butler J. Alveolar gas compression in smokers and asthmatics. Am Rev Respir Dis 1980; 121:291–295.
33. Miller A. A simple spirometric clue to asthma: airways obstruction suggested by negative or reduced forced expiratory reserve volume despite normal FEV$_1$–FVC ratio. Mt Sinai J Med 1990; 57:85–92.
34. American Thoracic Society. Definitions and classifications of chronic bronchitis, asthma and pulmonary emphysema. Am Rev Respir Dis 1962; 85:762–768.
35. Pennock BE, Rogers RM, McCaffree DB. Changes in measured spirometric indices: what is significant? Chest 1981; 80:97–99.
36. Sourk RL, Nugent KM. Bronchodilator testing: confidence intervals derived from placebo inhalations. Am Rev Respir Dis 1983; 128:153–157.
37. Miller A. Bronchodilator and provocational tests. In: Miller A, ed. Pulmonary function tests in clinical and occupational lung disease. Orlando: Grune & Stratton, 1986:307–323.
38. Ramsdell JW, Tisi GM. Determination of bronchodilation in the clinical pulmonary function laboratory. Chest 1979; 76:622–628.
39. Pare PD, Lawson LM, Brooks LA. Pattern of response to inhaled bronchodilators in asthmatics. Am Rev Resp Dis 1983; 127:680–685.
40. Cockcroft DW, Berocheid BA. Volume adjustment of maximal midexpiratory flow. Importance of changes in total lung capacity. Chest 1980; 78:595–600.
41. Light RW, Conrad SA, George RB. The one best test for evaluating the effects of bronchodilator therapy. Chest 1977; 72:512–516.
42. Juniper EF, Frith PA, Dunnett C, Cockcroft DW, Hargreave FE. Reproducibility and comparison of responses to inhaled histamine and methacholine. Thorax 1978; 33:705–710.

43. Bascom R, Bleecker ER. Bronchoconstriction induced by distilled water: sensitivity in asthmatics and relationship to exercise-induced bronchospasm. Am Rev Respir Dis 1986; 134:248–253.
44. Heaton RW, Henderson AF, Costello JF. Cold air as a bronchial provocation technique: reproducibility and comparison with histamine and methacholine inhalation. Chest 1984; 86:810–814.
45. Chatham M, Bleecker ER, Smith PL, Rosenthal RR, Mason P, Norman PS. A comparison of histamine, methacholine and exercise airway reactivity in normal and asthmatic subjects. Am Rev Respir Dis 1982; 126:235–240.
46. Pepys J, Hutchcroft BJ. Bronchial provocation tests in etiologic diagnosis and analysis of asthma. Am Rev Respir Dis 1975; 112:829–859.
47. Weiss JW, Drazen JM, McFadden ER Jr, Weller P, Corey EJ, Lewis RA, Austen KF. Airway constriction in normal humans produced by inhalation of leukotriene D. JAMA 1983; 249:2814–2817.
48. Chai H, Farr RS, Froehlich LA, et al. Standardization of bronchial challenge procedures. J Allerg Clin Immunol 1975; 56:323–327.
49. Ryan G, Dolovich MB, Roberts RS, Frith PA, Juniper EF, Hargreave FE, Newhouse MT. Standardization of inhalation provocation tests: two techniques of aerosol generation and inhalation compared. Am Rev Respir Dis 1981; 123:195–199.
50. Corrao WM, Braman SS, Irwin RS. Chronic cough as the sole presenting manifestation of bronchial asthma. N Engl J Med 1979; 300:633–637.
51. Malo J-L, Pineau L, Cartier A, Martin RR. Reference values of the provocative concentrations of methacholine that cause 6% and 20% changes in forced expiratory volume in one second in a normal population. Am Rev Respir Dis 1983; 128:8–11.
52. Pattemore PK, Asher MI, Harrison AC, Mitchell EA, Rea HH, Stewart AW. The interrelationship among bronchial hyperresponsiveness, the diagnosis of asthma, and asthma symptoms. Am Rev Respir Dis 1990; 142:549–554.
53. Kelso JM, Enright PL, Scanlon PD, O'Connell EJ, Sachs MI. Effect of inhaled methacholine on inspiratory flow. Chest 1990; 98:1426–1429.
54. Gluck EH, Onorato DJ, Castriotta R. Helium-oxygen mixtures in intubated patients with status asthmaticus and respiratory acidosis. Chest 1990; 98:693–698.
55. Schilder DP, Roberts A, Fry DL. Effect of gas density and viscosity on the maximal expiratory flow-volume relationship. J Clin Invest 1963; 42:1705–1713.
56. Despas PJ, Leroux M, Macklem PT. Site of airway obstruction in asthma as determined by measuring maximal expiratory flow breathing air and a helium-oxygen mixture. J Clin Invest 1972; 51:3235–3243.
57. Dosman J, Bode F, Urbanetti J, Martin R, Macklem PT. The use of a helium-oxygen mixture during maximum expiratory flow to demonstrate obstruction in small airways in smokers. J Clin Invest 1975; 55:1090–1099.
58. Ingram RH Jr, Wellman JJ, McFadden ER Jr, Mead J. Relative contributions of large and small airways to flow limitation in normal subjects before and after atropine and isoproterenol. J Clin Invest 1977; 59:695–703.
59. Pichurko BM, Ingram RH Jr, Sperling RI, Lafleur JE, Corey EJ, Austen KF, Drazen JM. Localization of the site of the bronchoconstrictor effects of leukotriene $C_4$ compared with that of histamine in asthmatic subjects. Am Rev Respir Dis 1989; 140:334–339.
60. Eliasson O, Zuwallack RL. Density dependence of maximal expiratory air flow in asthmatics with exacerbation of their disease. Am Rev Respir Dis 1987; 135:573–578.
61. Redline S, Tager IB, Castile RG, Weiss ST, Barr M, Speizer FE. Assessment of the usefulness of helium-oxygen maximal expiratory flow curves in epidemiologic studies of lung disease in children. Am Rev Respir Dis 1987; 136:834–840.
62. DuBois AB, Botelho SY, Comroe JH Jr. A new method for measuring airway resistance in man using a body plethysmograph: values in normal subjects and in patients with respiratory disease. J Clin Invest 1956; 35:327–335.
63. Takishima T, Hida W, Sasaki H, Suzuki S, Sasaki T. Direct-writing recorder of the dose-response curves of the airway to methacholine: clinical application. Chest 1981; 80:600–606.
64. Iwamoto I, Tanaka M, Ra C, Tomioka H, Yoshida S, Fujita Y, Inoue K. Differentiation in dose-response curves of airway responsiveness to inhaled methacholine between asthmatics and nonasthmatics. J Asthma 1989; 26:291–298.
65. Gayrard P, Orehek J, Grimaud C, Charpin J. Bronchoconstriction effects of a deep inspiration in patients with asthma. Am Rev Respir Dis 1975; 111:433–439.
66. Fish JE, Ankin MG, Kelly JF, Peterman VI. Regulation of bronchomotor tone by lung inflation in asthmatic and nonasthmatic subjects. J Appl Physiol 1981; 42:1079–1086.
67. Finucane KE, Colebatch HJH. Elastic behavior of the lung in patients with airway obstruction. J Appl Physiol 1973; 34:143–153.
68. Gold WM, Kaufman HS, Nadel JA. Elastic recoil of the lungs in chronic asthmatic patients before and after therapy. J Appl Physiol 1967; 23:433–438.
69. Otis AB, McKerrow CB, Bartlett RA, Mead J, McIlroy MB, Selverstone NJ, Radford EP. Mechanical factors in distribution of pulmonary ventilation. J Appl Physiol 1956; 8:427–443.
70. McFadden ER Jr, Kiker R, Holmes B, de Groot WJ. Small airway disease: an assessment of the tests of peripheral airway function. Am J Med 1974; 57:171–182.
71. Deychakiwsky YA, Deal EC Jr, Saidel GM. Ventilatory inhomogeneity associated with acute bronchoconstriction in asthmatic patients. Respiration 1985; 47:201–208.
72. Wanner A, Zarzecki S, Atkins N, Zapata A, Sackner MA. Relationship between frequency dependence of lung compliance and distribution of ventilation. J Clin Invest 1974; 54:1200–1213.
73. Burns CB, Scheinhorn DJ. Evaluation of single-breath helium dilution total lung capacity in obstructive lung disease. Am Rev Respir Dis 1984; 130:580–583.
74. Brown R, Ingram RH Jr, McFadden ER Jr. Problems in the plethysmographic assessment of changes in total lung capacity in asthma. Am Rev Respir Dis 1978; 118:685–692.
75. Stanescu DC, Rodenstein D, Cauberghs M, Van de Woestijne KP. Failure of body plethysmography in bronchial asthma. J Appl Physiol 1982; 52:939–948.
76. Rodenstein DO, Stanescu DC, Francis C. Demonstration of failure of body plethysmography in airway obstruction. J Appl Physiol 1982; 52:949–954.
77. Muller N, Bryan AC, Zamel N. Tonic inspiratory muscle

activity as a cause of hyperinflation in histamine-induced asthma. J Appl Physiol 1980; 49:869–874.

78. Martin J, Powell E, Shore S, Emrich J, Engel LA. The role of respiratory muscles in the hyperinflation of bronchial asthma. Am Rev Respir Dis 1980; 121:441–447.
79. Wagner PD, Dantzker DR, Iacovoni VE, Tomlin WC, West JB. Ventilation-perfusion inequality in asymptomatic asthma. Am Rev Respir Dis 1978; 118:511–524.
80. McFadden ER Jr, Lyons HA. Arterial-blood gas tension in asthma. N Engl J Med 1968; 278:1027–1032.
81. Cosgrove JF, Neuberger N, Levison H. The ventilatory response to carbon dioxide in asthmatic children, measured by the mouth-occlusion method ($P_{M100}$). Pediatrics 1976; 57:952–959.
82. Savoy J, Fleetham JA, Arnup ME, Anthonisen NR. Airway anesthesia and respiratory response to methylcholine induced bronchoconstriction. Respir Physiol 1981; 43:59–68.
83. Ohman JL Jr, Schmidt-Nowara W, Lawrence M, Kazemi H, Lowell FC. The diffusing capacity in asthma: effect of airflow obstruction. Am Rev Respir Dis 1973; 107:932–939.
84. Weitzman RH, Wilson AF. Diffusing capacity and over-all ventilation: perfusion in asthma. Am J Med 1974; 57:767–774.

# Bone Marrow Failure Disorders

Blanche P. Alter, M.D.

---

I discuss here a variety of aplastic anemias, both acquired and inherited, a very broad topic (for detailed references see [1] and [2]), and I start with the punchline: The hypothesis is that what we refer to as acquired aplastic anemia appears on the background of homozygosity, or at least heterozygosity, for genes for bone marrow failure. I discuss first those disorders that appear to be acquired aplastic anemias, entities to which we can associate some kind of causality, then discuss a few of the many inherited bone marrow failure disorders, pointing out that in few can we actually identify patients who do have genes for aplastic anemia based on their being homozygous for recessive diseases or having an X-linked disorder. Heterozygotes, who are the vast majority, cannot be diagnosed at this time. In probably less than five years, however, at least some of the genes will have been identified, and then this whole hypothesis can be tested.

To provide a context: The number of cases of "acquired aplastic anemia" in the United States per year is between two and five per million population. The heterozygote frequency for one of the inherited aplastic anemias, Fanconi's anemia (see below), is about 1 in 200. So every single patient with acquired aplastic anemia could, in fact, be a heterozygote for Fanconi's anemia.

**Causation.** Among the possible causes of acquired and inherited aplastic anemia, the "idiopathic" entity accounts for at least 50% of the cases. This can be construed in two ways. First, in many patients we have not yet identified the cause—possibly because we did not take a good enough history. The other possibility, of course, is that we are assigning causes incorrectly but by coincidence. Perhaps both suggestions are true.

The secondary causes of aplastic anemias include

- radiation, such as an atomic bomb blast or a radiation plant problem, as occurred in Chernobyl
- exposure to certain drugs and chemicals, regularly following exposure to cytotoxic chemotherapy, and also to large enough amounts of organic solvents such as benzene; idiosyncratically following exposure to certain other drugs—chloramphenicol, some of the nonsteroidal anti-inflammatories, antiepileptics
- gold, which is used therapeutically
- several viruses: Epstein-Barr virus, associated with mononucleosis; hepatitis (more than 100 cases of aplastic anemia have been related to this virus); parvovirus; human immunodeficiency virus
- immune disorders (eosinophilic fasciitis, hypoimmunoglobulinemia)
- thymoma
- changes during pregnancy; several pregnant women have developed aplastic anemia, and in several more pregnant aplastic women the aplasia got worse, suggesting a relationship with estrogens
- paroxysmal nocturnal hemoglobinuria
- some of the preleukemic syndromes

Equally important, at least to me, are the inherited aplastic anemias. Although inherited, they are not exclusively disorders of infants and children, since some are not diagnosed until the patients are adults, and many of the patients live long enough to be seen by internists.

---

Adapted from the author's Grand Rounds in Medicine presentation April 3, 1990, at The Mount Sinai Medical Center.

Supported in part by National Institutes of Health grant HL26132, the National Organization for Rare Disorders, and the Fanconi Anemia Research Fund.

From the Polly Annenberg Levee Hematology Center, Departments of Medicine and Pediatrics, The Mount Sinai Medical Center, New York, New York. Address reprint requests to Blanche P. Alter, M.D., Box 1079, Division of Hematology, Mount Sinai Medical Center, One Gustave L. Levy Place, New York, NY 10029.

These inherited causes include

- Fanconi's anemia, an autosomal recessive condition; more than 700 cases are reported in the literature, and more than 200 are known in the United States, some with physical anomalies
- dyskeratosis congenita, primarily but not exclusively an X-linked disorder, with abnormal dermal manifestations; aplastic anemia occurs in 50% of the cases
- Shwachman-Diamond syndrome, pancreatic insufficiency with neutropenia; about 20% of persons with this syndrome go on to pancytopenia
- reticular dysgenesis, lack of all lineages at birth
- amegakaryocytic thrombocytopenia, thrombocytopenia at birth; most of these disorders evolve into aplastic anemia
- familial aplastic anemias; among families in which an acquired aplastic anemia has occurred are eight families in Scandinavia with 19 members who have aplastic anemia, again supporting a genetic propensity (3)
- preleukemias, myelodysplasia, monosomy 7
- nonhematologic syndromes; for example in Down's, Dubowitz, and Seckel syndromes, aplasia has occurred in unexpectedly high numbers

There are also the single cytopenias, both acquired and inherited (Table 1). On an immune basis, thymoma has been associated with pure red cell aplasia. The pediatric equivalent of pure red cell aplasia, transient erythroblastopenia of childhood, also often has an immune basis but may be virally related. The inherited disorders include Diamond-Blackfan anemia, which is red cell aplasia that usually remains a single cytopenia; and Kostmann's syndrome, which is genetic agranulocytosis—the only disorder on this list that can now be treated with a growth factor. These patients have neutropenia that is improved with infusions of granulocyte colony-stimulating factor. Shwachman-Diamond syndrome, reticular dysgenesis, and amegakaryocytic thrombocytopenia may go on to aplasia. Thrombocytopenia absent radii is a single cytopenia. The thrombocytopenia often improves as the patients get older, and to my knowledge, they have never developed aplastic anemia or leukemia.

TABLE 1
*Classification of the Single Cytopenias*

| Anemia | Neutropenia | Thrombocytopenia |
|---|---|---|
| Acquired | | |
| Idiopathic | Idiopathic | Idiopathic |
| Drugs, toxins | Drugs, toxins | Drugs, toxins |
| Immune | | |
| Thymoma | | |
| TEC | | |
| Inherited | | |
| Diamond-Blackfan | Kostmann's<br>Shwachman-Diamond<br>Reticular dysgenesis | Amegakaryocytic<br>Thrombocytopenia absent radii |

TEC, transient erythroblastopenia of childhood.
Modified from Ref. 1 in Alter and Young (2).

## Acquired Aplastic Anemia

**Criteria for Diagnosis.** The definition of aplastic anemia is fairly clear but must be specific and requires the presence of peripheral pancytopenia with a hypocellular bone marrow. The marrow diagnosis requires both an aspiration for cellular analysis and a biopsy for overall cellularity. There must be distinction between patients whose pancytopenia is due to hypersplenism and destruction of their cells and patients who have marrow infiltration, collagen vascular disease, or hypercellular bone marrows. This distinction

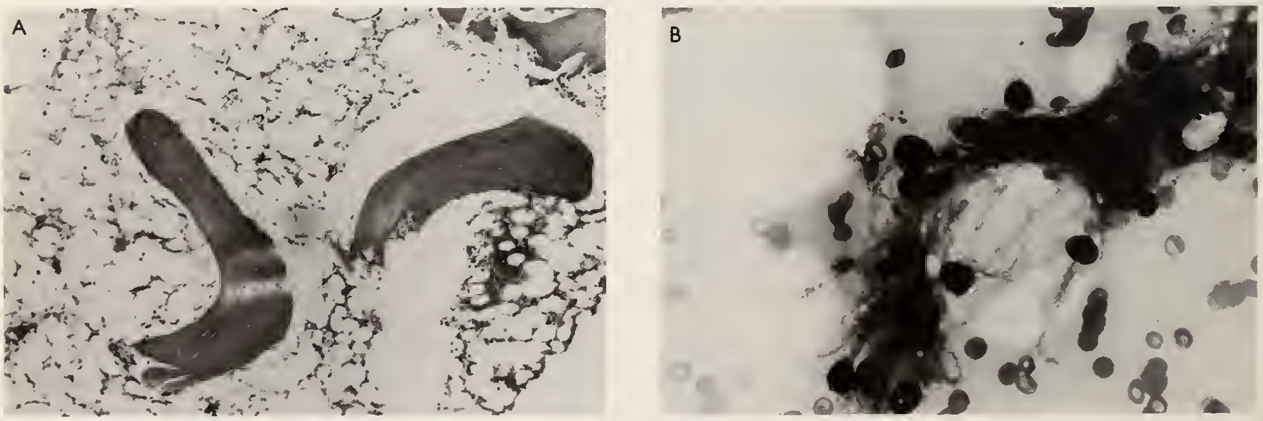


Fig. 1. Bone marrow examination in severe aplastic anemia. (A) Biopsy; note hypocellularity. (B) Aspirate; residual cells are lymphocytes, stromal cells, plasma cells, and mast cells. *Courtesy of* Gail Wolfe, MD. From Alter (1).

must be made because the choice of therapy depends on the diagnosis.

Most discussion of aplastic anemia refers to the severe condition because the outcome is quite different in patients who fulfill these criteria than in patients who have a moderate condition who have some residual hematopoietic progenitors in their bone marrow. Severe aplasia was defined by a collaborative marrow transplant group many years ago in a paper by Camitta et al. (4). The absolute neutrophil count is under 500/fL, platelets below 20,000/fL, and reticulocytes are less than 1% when corrected for hematocrit (many of these patients, of course, are transfused by the time they are seen by the hematologist). The marrow shows severe or moderate hypocellularity, but less than 30% of the residual cells in the marrow are hematopoietic precursors.

Even physicians who are not accustomed to looking at marrow biopsies might recognize aplasia (Fig. 1). The biopsy shows the bony cortex, small amount of bone marrow, and large amount of fat that is clearly excessive, regardless of the age of the patient. In addition, most of the cells are rather monotonous and most are lymphocytes. There are very few residual myeloid and erythroid elements.

**Relation to Age.** One of the issues to address is the relationship of aplasia to age. There are several ways to approach this, none totally successful, because without an international registry or mandatory reporting, the relationship is impossible to determine. Data taken from death certificates reported to the U.S. Public Health Service show that the age relationship remains fairly steady through the early part of life; in older age groups, the incidence becomes extremely high. There are several possible explanations. One is that humans have a fixed number of hematopoietic progenitor cells, with a finite number of divisions, that are eventually going to run out, leading to aplasia. Another possibility is that the older individual has a history of cumulative exposure to all the noxious agents that can cause aplasia and that bone marrow finally gives out. Aplasia in older individuals may be a combination of all these.

Data on age and aplastic anemia are also available from the Seattle tertiary referral center for marrow transplant, where most patients older than 30 years of age, and all of those older than 50 years, are not referred for transplant. These data are biased toward the younger group and show a median of around 20 years of age. Similar data were provided by Neal Young at the National Institutes of Health, which does not perform transplants but does run a referral service for treatment with antithymocyte globulin and also cyclosporin A. Although these data should not reflect the age restriction that data based on transplant subjects would generate, but does reflect a bias based on referrals, the peak age is also around 20 years of age.

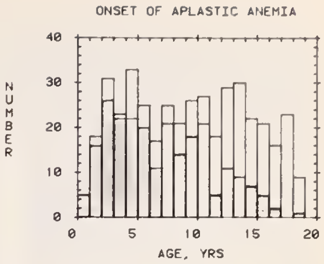


Fig. 2. Age at diagnosis of aplastic anemia in patients under age 20 reported in the literature. Gray area is from pediatric, white area from adult literature reports. From Alter (1).

I am a pediatrician, and I have written a textbook chapter on aplastic anemia in a pediatric hematology text for a number of years (1, 2). The first time I wrote the chapter, I read the pediatric literature and assembled the data shown in Fig. 2 to examine the age relationship in children to age 20. There seemed to be a peak in the 5- to 10-year age range (gray area in figure). However, I started reading more broadly on aplasia and have tabulated all cases, including those outside the pediatric literature, in which the patients were under age 20. In my tabulation the age incidence is in fact quite even, so all the available age data depend on how you look at it, and probably all are incorrect. But they are all we have.

## Mechanisms

In the hematopoietic system (Fig. 3) the marrow is derived from a pluripotent stem cell, which has not been isolated in humans, although Irving

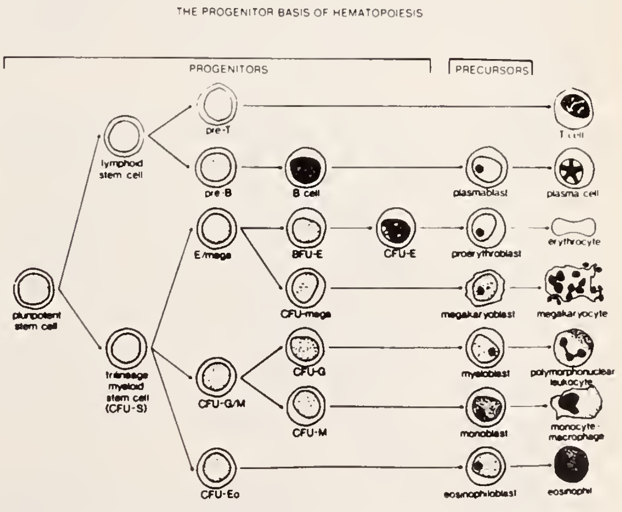


Fig. 3. Progenitor basis of hematopoiesis, showing progenitor and precursor compartments. From Nathan (5).

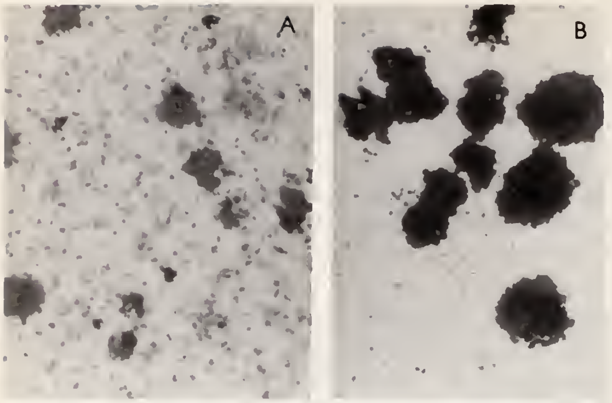


Fig. 4. (A) Colonies derived from CFU-E; (B) burst derived from BFU-E.

Weissman at Stanford apparently has done so in mice (6). This cell presumably gives rise to the lymphoid and the trilineage myeloid lineages. Peripheral blood contains red cells, white cells, and platelets, whereas bone marrow contains precursors, which are morphologically recognizable cells: erythroblasts, myeloblasts, promyelocytes, and so forth. The progenitor compartment is essentially a black box. All these cells are mononuclear cells that look like lymphocytes. They are identified by functional assays, cultures in semisolid medium with the appropriate stimulating factors.

For example, to examine the erythroid lineage, which is what my laboratory does, the hormone erythropoietin is added to the cultures. Examination of bone marrow cultures (Fig. 4) at 7 days shows small colonies of perhaps 32 to 64 erythroblasts each, derived from an erythroid colony-forming unit (CFU-E). Examination at 14 days shows larger colonies derived from the more primitive erythroid progenitor cell called the burst-forming unit erythroid (BFU-E). There are similar colony-forming units for all the other lineages.

Aplastic anemia could therefore be due to an abnormality in the pluripotent stem cell or in the trilineage stem cell. In the mouse the latter has been called the colony-forming unit spleen, but that is not an assay we can do in humans. Aplasia in the single cytopenias may be due to an inability to differentiate along one of the specific lines, or to a lack of one of the factors that is relevant to the maturation of these cell lines. Aplasia may in some patients be due to a defect in the stromal cell, in the microenvironment in which all hematopoiesis occurs.

What does this all mean for aplastic anemia? The multipotent progenitor cell undergoes both self-renewal and terminal differentiation following divisions. The whole hematopoietic compartment can be amplified because of this process of division followed by terminal differentiation, but eventually the cells stop dividing. They thus lose their proliferative potential as they gain in luxury functions and maturation. This can be envisioned as a pyramid, where a few self-renewing cells repopulate or populate the entire marrow, but eventually cells leave the bottom of the pyramid by terminal differentiation. This is death by differentiation, if you will. In aplasia, one possibility is that there is an intrinsic or exogenously induced damage to the DNA of the cells. An example is Fanconi's anemia, a chromosome breakage syndrome, in which the DNA is intrinsically abnormal. Other, external, agents may also damage DNA. When this happens, the more mature cells that are not dividing may leave the hematopoietic compartment, but they may not be replaced if the earliest cells are too damaged and too few. On the other hand, if enough intact residual cells remain, the patient may recover after alleviation of the factor that caused the aplasia. Alternatively, there are agents that may damage all cells independent of whether they are dividing. If all the cells are damaged, they will leave the hematopoietic compartment, but there will be nothing to replace them. If there are enough intact residual cells in the self-renewing repopulation compartment, the patient can recover.

**Radiation.** One cause of aplastic anemia is radiation. The usual measurements that physicians deal with, such as blood volume, red-cell life spans, pharmacokinetics, Shilling tests, computed-tomography scans, and natural radiation, all entail lower levels of radiation. Some radionuclide scans are a little higher. The ordinary or even diagnostic levels of radiation are not going to cause aplasia, which requires levels such as occurred at Hiroshima and Nagasaki, or in Chernobyl.

**Viruses.** Viruses may be the cause of aplastic anemia.

B19 parvovirus is a single-stranded DNA virus, with a somewhat specific target in erythroid progenitor cells. It is the cause of aplastic or hypoplastic crises in patients with hemolytic anemia. Most of these are patients with sickle-cell disease, but parvovirus can be relevant in any hemolytic anemia. Parvovirus also explains why patients with sickle-cell disease usually have only one such episode in their history. Pure red-cell aplasia has occurred in patients with immunodeficiency who were unable to mount an antibody to the virus. And probably between 5% and 20% of fetuses whose mothers have a parvovirus

**TABLE 2**
*Drugs and Chemicals as Causes of Aplastic Anemia, 1950–1989*

| Agent | 1950–1959 (%) | 1960–1969 (%) | 1970–1979 (%) | 1980–1989 (%) | Far East 1970–1979 (%) |
|---|---|---|---|---|---|
| Chloramphenicol | 30 | 27 | 17 | 3 | 19 |
| Phenylbutazone | 1 | 9 | 6 | 3 | 2 |
| Sulfonamides | 4 | 1 | 1 | 1 | 1 |
| Anticonvulsants | 2 | 2 | 2 | 1 | 1 |
| Gold | 0 | 1 | 1 | 2 | <1 |
| Benzene, solvents | 6 | 4 | 3 | 3 | 3 |
| Insecticides | 1 | 9 | 1 | 1 | 4 |
| Other drugs | 16 | 8 | 13 | 7 | 12 |
| Total drugs | 57 | 60 | 45 | 20 | 43 |
| Total no. of cases | 756 | 339 | 1292 | 811 | 3391 |

Modified from Ref. 1 in Alter and Young (2).

infection are at risk for severe hydrops fetalis. Parvovirus was recently identified as the cause of fifth disease in children. The childhood rashes were numbered one through six, and included measles (rubeola), German measles (rubella), and scarlet fever. The cause of the fifth, called erythema infectiosum, has now been identified as parvovirus. In children, it causes a viral syndrome entailing a reddish face and high fever. In adults, this same viral illness is associated with joint symptoms about a week later due to immune complex formation.

Other viruses that cause aplastic anemia include cytomegalovirus, which is sometimes the cause of bone marrow transplant graft failure; infects fibroblasts—that is, the stroma; and perhaps also infects progenitors. Epstein-Barr virus clearly infects lymphocytes and macrophages and perhaps erythroid progenitors; it causes infectious mononucleosis, and sometimes aplasia is associated with it. The flavoviruses, such as dengue and hepatitis, particularly hepatitis C, involve macrophages and perhaps progenitors. Human immunodeficiency virus clearly involves macrophages, although not so clear is whether it actually infects progenitors as well.

There are several mechanisms for viral aplasia. Parvovirus infects the erythroid progenitor cell, replicates within it, and causes that cell to lyse. Epstein-Barr virus infects a cell but does not kill it, but does change its antigenic characteristics such that activated T cells now recognize that these cells are abnormal and destroy them. Viruses may inhibit luxury functions by replicating within cells and preventing proliferation or maturation. A virus such as cytomegalovirus may infect stromal cells, which may be the source of hematopoietic growth factors that cause the hematopoietic stem cell to grow. If the stromal cell is either killed or impaired in its function, the growth factors are not produced and the cells cannot be made.

**TABLE 3**
*Characteristics of Two Types of Hematologic Toxicity from Chloramphenicol*

| | Reversible toxicity | Irreversible aplasia |
|---|---|---|
| Incidence | Common | Rare |
| Bone marrow cellularity | Cellular | Aplastic |
| Relation to dose | Dose dependent | Unrelated to dose |
| Onset | Concurrent with drug therapy | 3 wk–5 mo from last dose |
| Bone marrow changes | Maturation arrest, erythroid; cytoplasmic or nuclear vacuolization | Aplasia |
| Peripheral blood | Reticulocytopenia, occasional leukopenia, thrombocytopenia; anemia if treatment prolonged; elevated serum iron | Pancytopenia |
| Initial symptoms | None; pallor due to anemia (if treatment prolonged) | Bleeding, infection |
| Outcome | Reversible | Usually fatal |

From Ref. 8 in Alter (1).

In the normal individual, parvovirus infects the erythron. The virus is then released as the cells are destroyed, antibody is made, immune complexes are formed, and the pattern of fifth disease appears. The patient with a hemolytic anemia and an expanded erythron has virus excess because many erythroid cells are being destroyed. These patients do not mount an antibody response fast enough and thus have a transient aplastic crisis, which occurs only in individuals with shortened red cell survival. Although the reticulocyte count may decrease in hematologically normal individuals, they can survive a week or two without having made red cells, but patients with hemolytic anemia may not be able to survive.

The erythron of the patient who has an immune disorder is infected by the virus, virus is released, but no antibody is formed. The virus goes back and infects more erythroid cells and the erythron shrinks. This cycle can only be broken by providing immunoglobulin. One such patient, who was reported in the literature as having familial pure red cell aplasia, actually had familial immunodeficiency (7). This individual was anemic, had no reticulocytes, but was found, using modern techniques of gene amplification, polymerase chain reaction amplification of DNA in the serum, to have parvoviral DNA. After being treated with immunoglobulin, he began to have reticulocytes and an increase in hemoglobin. Ultimately phlebotomy was needed to remove all the iron that had accumulated during the years of transfusion.

**Drugs and Chemical Agents.** At the top of the list of drugs and chemicals that can cause aplasia is chloramphenicol (Table 2). In the 1950s, when this drug was used rather widely, it was responsible for 30% of reported cases of aplastic anemia. It has been noted that a large number of aplastic anemia cases occur in the Far East, where many of the drugs are available over-the-counter. However, separate analysis of cases from the Far East showed that the proportion of drug-induced cases was actually the same as in the Western world; hepatitis-induced aplasia may be more common in the Orient. In any event, if we examine the most recent decade, the proportion of aplasias caused by drugs is substantially reduced. One reason for that is that physicians are more careful in treating and monitoring their patients because they are aware of the associations of drugs and chemicals with aplasia. The second cause may be underreporting, because this is phenomenology, and physicians are not as interested in that anymore. The interest is in how we can cure patients, and whether growth factors are going to work or whether marrow transplant can be used.

TABLE 4
*Immune Pathophysiology in Aplastic Anemia*

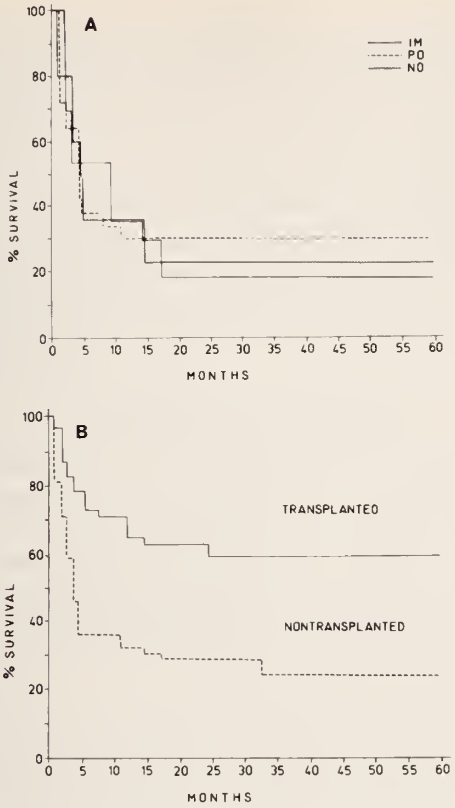
| Phenomenon | Positive studies | Negative studies |
|---|---|---|
| Activated T cells | CD8+, DR+, IL-2R+ | No DR+ |
| | TAC+ pre-BMT | Normal |
| | T8+ suppression | Normal |
| γ-Interferon | Spontaneous or stimulated production | Normal |
| γ-Interferon in serum | Bioassay or RIA | Normal |
| IL-2 production | Increased | Normal |
| CD4+:CD8+ | Reversed | Normal |

See Alter and Young (2) for references.

Drug mechanisms can be discussed briefly. Drugs or chemicals may be specifically toxic to the hematopoietic cell. The individual may be unable to detoxify an intermediate product that was made in the body using the drug owing to an inability to acetylate or methylate or owing to specific genetic defects in the cytochrome P-450 enzyme system. Toxic intermediates may be made that the patients cannot detoxify, or drugs may be actively converted to toxic intermediates that cannot be removed.

The antibiotic chloramphenicol warrants further discussion (Table 3). I will date myself and perhaps a few others when I say that as an intern, I learned that treatment of meningitis was with penicillin and chloramphenicol. At that time we were certainly aware of the reversible marrow suppression from chloramphenicol. As a medical student and a house officer, I was doing reticulocyte counts daily, because the hospital laboratories did not provide that service. A fall in reticulocyte count is one of the early signs of toxicity to this drug. The marrow shows a maturation arrest, which occurs while the patient is taking the drug, is dependent on the dose, has a higher risk in patients with liver disease who may not adequately glucuronidate the chloramphenicol, and also occurs in patients with renal disease because they may not excrete the chloramphenicol. Chloramphenicol-induced marrow suppression is usually reversible, but patients have died from sepsis or other complications while they were in the suppressed phase. So the dose-related marrow suppression is not totally benign.

The idiosyncratic or "allergic" type of chloramphenicol-induced aplasia is identified only by history; it occurs after the patient is no longer taking the drug. In the 1960s it was reported in 1 in 20,000 patients in the state of California who were given chloramphenicol for "viral" or cold-type symptoms. So it has a fairly high frequency. I do not think chloramphenicol is used for this type of indication anymore. The toxicity was usually severe aplasia that was usually fatal. Some



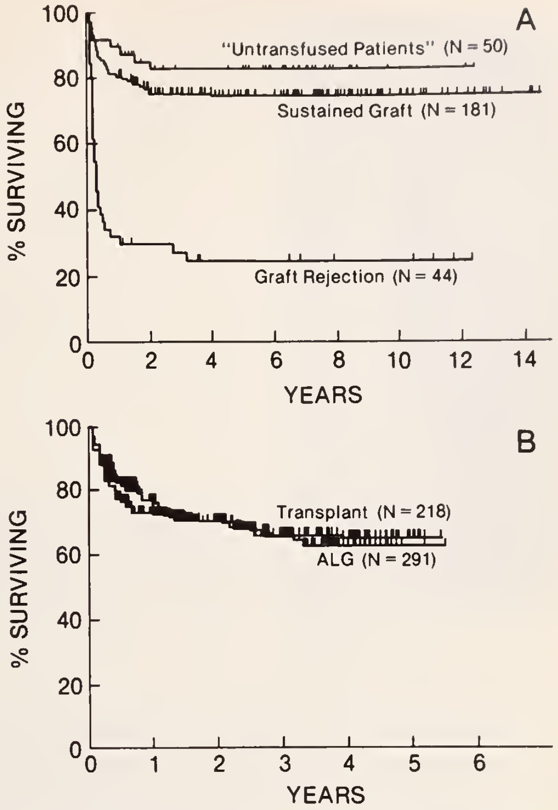


Fig. 5. **Left** Kaplan-Meier plots of cumulative survival in severe aplastic anemia. Time is shown as months from diagnosis. (A) Results of treatment with intramuscular (IM) or oral (PO) androgen compared with no treatment (NO). (B) Results of treatment with marrow transplantation. From Camitta et al. (10). Fig. 6. **Right** Actuarial survival in severe aplastic anemia. (A) Bone marrow transplantation, showing critical role of graft acceptance in determining outcome. From Deeg et al. (11). (B) European data on bone marrow transplantation vs. immunosuppression. From Anasetti et al. (11a) and Gordon-Smith (12).

other drugs can be classified in the same way, as either dose-related or idiosyncratic.

**Immune Basis.** Immune mechanisms have been invoked as causes for aplasia, and some studies have suggested activated T cells (Table 4); other studies do not agree. In some reports production of $\gamma$-interferon was thought to be increased; in others it was thought to be normal. Interleukin 2 was also found to be increased in some, and normal in others. Reversal of the CD4:CD8 ratio was noted in only some studies. I think these unclear results reflect the heterogeneity of the patients. Each study is a small one, since no one can gather a huge series of patients, all with the same kind of aplastic anemia. The unclear results also reflect the heterogeneity of the laboratories, because each laboratory does its own assay in its own way and may elicit different results. Suffice it to say that we really do not have good proof that aplasia is an immune disorder, except that patients can be cured with immune therapy. Thus there is some connection, but what the connection is, is not clear. It may just be an imbalance in marrow between the appropriate stimulatory and suppressor cells; that imbalance may be what immune therapy addresses.

## Treatment

There are several approaches to treatment of aplasia.

The primary consideration should be bone marrow transplantation. It is said that 25%–30% of the patients may be eligible, but in fact 20% or less actually do have an HLA-matched sibling donor and are younger than perhaps 30 or 40 years. Marrow transplant may cure the patient; mortality is probably 10%–25%.

For those who do not have a marrow transplant donor, every other possible treatment becomes important. At the moment the second choice is antithymocyte globulin (ATG). Although aplastic anemia may have something to do with the immune system, we are not entirely clear how ATG works. It is variable, however, because ATG is not always available and the lots that are available may not work well. The next approach after ATG, or when that fails, is cyclo-

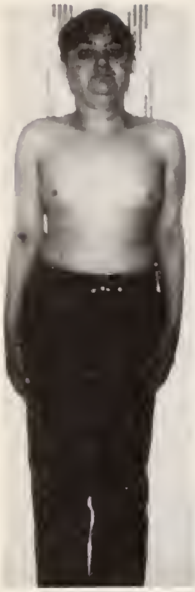


Fig. 7. A 37-year-old man with Fanconi's anemia. *Courtesy of* Arleen Auerbach, Ph.D.

sporin A, also an immune reagent whose specific role in aplasia is not clear.

The Europeans have had some success using high-dose methylprednisolone intravenously for about three weeks. That has not worked as well on this side of the Atlantic, for some reason. Androgens must also be considered, more for historical reasons than anything else. In fact, the first paper I wrote in hematology laid to rest the myth that androgens would cure AA (9). Probably most of the studies in which androgens were successful included mainly patients with moderate and not severe aplasia, that is, patients who had residual stem cells. In the inherited disorders, androgens become important in treatment.

Lower on the therapy list, but not in terms of interest, are the hematopoietic growth factors. Some ongoing studies are now using granulocyte macrophage colony-stimulating factor (GM-CSF). Studies that are just beginning are using interleukin 3 (IL-3), which acts perhaps a little earlier in the hematopoietic growth scheme than does GM-CSF, and some think it may be even more useful in aplastic anemia. Again, the initial reports are not totally encouraging, except in patients with moderate aplasia who have residual stem cells.

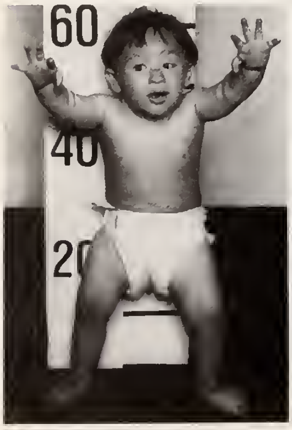


Fig. 8. A three-year-old boy with Fanconi's anemia. From Alter and Young (2).

Antiviral therapy may be important in those who have a viral disease. In addition, immunoglobulin therapy may be specific for those who have parvovirus infection and are immunodeficient. And finally, although not least important, is supportive care. The obvious needs repetition here: until a marrow transplant is ruled out, a family member must not be used as a blood product donor, to avoid sensitization. Although that seems obvious, recently a patient of mine with one of the inherited aplasias was en route to the hospital for a transfusion when the patient's mother called me and said her pediatrician had ordered that the HLA-matched brother be used as the blood donor, although he was to be the marrow donor.

In an early study of Camitta et al. (10), untransplanted aplastics, including androgen-treated patients identical to the nonandrogen-treated group, were compared with patients who had marrow transplants (Fig. 5). The 50% mortality in untransplanted patients who had severe aplastic anemia occurs within the first six months after diagnosis. Even projected survival curves in leukemia are better than this. The transplanted group at that time—the early 1980s—had approximately 70% survival, and of course it is better now. In more recent data from Seattle (11a), a group of patients who had no blood products had approximately a 90% survival rate following marrow transplant, and all comers, including those who had transfusions, had a survival rate close to 80%. Those who rejected their graft were very similar to those who never got a transplant, with a rapid early mortality rate. In European data from Gordon-Smith (12) comparing marrow transplant with antilymphocyte globulin (ALG), survival with either modality is 60%–70% (Fig. 6); this report actually suggests starting with ALG and transplanting only nonresponders. Our concern is that the patients will have received supportive blood products in order to receive ALG, and that may jeopardize the transplant. Thus this is not our approach if the patient does have a marrow donor.

A constant question is whether to use unrelated donors or what are called "mismatched" relatives—close relatives who may be mismatched at one or two loci. The number of such transplants reported is low, and the survival is less than 25%.

The results with phenotypically matched but unrelated donors are also poor, with a 20%–30% survival rate, which is really about the same as in untransplanted, well-supported aplasia. Part of the problem is that if no HLA-matched sibling exists, much time has been lost at the point a donor is finally identified; blood products have been given to the patient, leading to a higher rate of graft rejection. In addition, these donors are not perfectly matched, so there may be graft-versus-host disease.

### Inherited Aplastic Anemia

To turn now from the acquired aplastic anemias to the inherited disorders, and to put them into a context: In a 20-year interval at Boston Children's Hospital, there were 134 patients with aplastic anemia, probably under age 20 years, of whom 40 had some form of an inherited disorder. Most of these had Fanconi's anemia, some had familial but not Fanconi's anemia, and four had amegakaryocytic thrombocytopenia that went on to aplasia. These figures suggest that probably 30% of pediatric aplasia is inherited. In fact, this ratio is not restricted to infants and children. In an ongoing study of patients over 16 years of age with aplasia referred for ATG treatment to the National Institutes of Health (Neal Young, personal communication), 5%–10% were found to have Fanconi's anemia (tested by chromosome breakage, discussed below). Results of their physical examinations were normal. Those of us who learned in medical school that patients with Fanconi's anemia have physical anomalies must now learn that these patients may be physically normal (see below). As I stated at the beginning of this discussion, there are several genes for aplastic anemia, one of which we can identify in the homozygote state (Fanconi's anemia). This condition will not be diagnosed if it is not thought of.

What may be the oldest currently surviving patient with Fanconi's anemia (Fig. 7) is 4 feet 10 inches tall at age 37 years. He has small, close-set eyes, a little bruising, and a little gynecomastia, owing to his treatment with androgens and prednisone. His hand anomalies include absence of the thenar muscles (the muscle at the base of the thumb), underdevelopment of thumbs, and a fixed joint on the left side, as well as small fifth fingers. He also has several café au lait spots.

I have other adult patients who have few physical anomalies. For example, there were two sisters, both approximately 5 feet tall, with skin darkly pigmented as if from a perpetual suntan. Both had café au lait spots, seen only when they were examined completely. This family was reported in 1961 as "pre-anemic Fanconi's anemia" (13). An older sister had physical anomalies and died of aplastic anemia, in the days before chromosome breakage analysis. The two sisters whom I saw did have chromosome breakage and had Fanconi's anemia. The older of the two died recently with a preleukemic refractory anemia. She also had a basal cell carcinoma and a cervical carcinoma in situ. The remaining affected sister, now 34 years old, does not like to visit doctors, but I believe she is doing well at the moment.

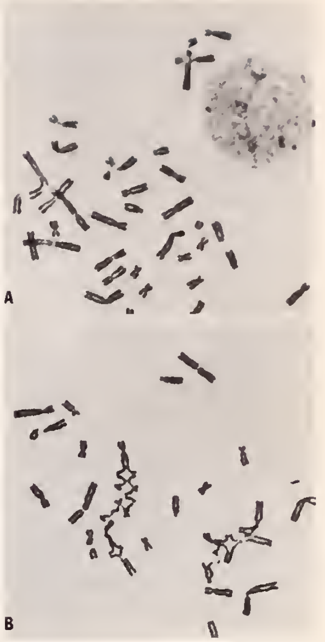


Fig. 9. Cytogenetic findings in Fanconi's anemia. (A) Baseline pattern. (B) Following culture with diepoxybutane. From Auerbach et al. (14).

Some patients are truly "textbook cases." One three-year-old boy (Fig. 8) has microcephaly, microphthalmia, a broad nasal base, epicanthal folds, and pseudostrabismus because of the epicanthus. He is short, has congenital hip dislocations, and his thumbs hang by a thread. His radii are hypoplastic and his feet rocker bottom, and he has café au lait spots and hypopigmented areas. He also had an imperforate anus, and his ureters were reimplanted because of reflux. He does not as yet have aplastic anemia. This type of patient constituted the majority until the chromosome breakage assay was developed (see below).

Approximately 700 patients with Fanconi's anemia are reported in the literature. The condition is clearly autosomal recessive. The mean age at diagnosis of aplasia is around age eight, and at least 10% of the patients are older than 16 at the time of diagnosis. Since this analysis does not include the patients that Young is ascertaining from the National Institutes of Health, the number of older patients will increase. And that may

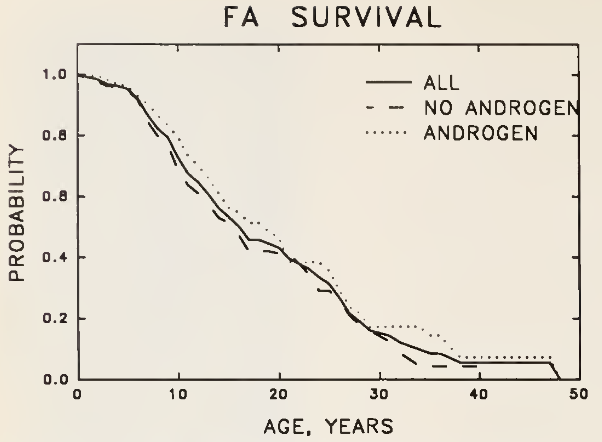


Fig. 10. Kaplan-Meier plot of cumulative survival in Fanconi's anemia (FA); time shown as years since diagnosis. Straight line (All), 678 cases reported from 1927 to 1990; broken line (No Androgen), 440 patients who apparently did not receive androgens; dotted line (Androgen), 238 patients who did receive androgens. From Alter and Young (2).

be the tip of the iceberg. People other than pediatricians do not tend to think of Fanconi's anemia as a cause of aplasia. And if the physical examination results are normal, nobody—neither pediatricians nor internists—tends to think of it.

The specific chromosome breakage analysis is done in the following way (14, 15): Peripheral blood is obtained, not bone marrow. It is cultured for three days so that cells will divide in the laboratory and form mitotic figures. For the last 24 hr, a DNA cross-linking agent such as diepoxybutane or mitomycin C is added to the culture. Patients with Fanconi's anemia are unable to repair cross-link damage, so their mitotic figures following this treatment are highly unusual (Fig. 9). The reason blood rather than bone marrow is tested is that in most laboratories, bone marrow chromosomes are examined immediately and not cultured. The breakage may be increased spontaneously but it is greatly magnified with these agents. So these specific tests are necessary. The other chromosome breakage syndromes, Bloom's syndrome, xeroderma pigmentosa, and ataxia telangiectasia, share with Fanconi's anemia an increase in spontaneous chromosome breakage, but they do not show a further increase with DNA cross-linking agents, which is specific for Fanconi's anemia.

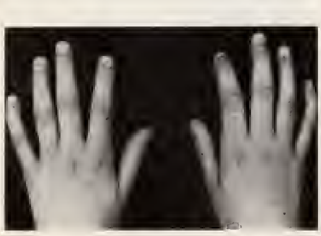


Fig. 11. Hands of patient with dyskeratosis congenita, demonstrating changes in fingernails. From Alter (1).

## Treatment

Although androgens do not work in aplastic anemia unless there are residual stem cells, the first line of treatment of Fanconi's anemia is androgens, if they can be used before the patient is totally aplastic. We usually use an oral androgen just because it is easier to manage, particularly in thrombocytopenic patients. It does carry a risk of liver toxicity, and we therefore monitor liver function carefully. Androgens are combined with small amounts of corticosteroid for a couple of reasons: androgens will accelerate growth, and steroids may delay it (these properties are relevant only when the patient is an infant or a child); and patients receiving steroids may bleed less at a given platelet count, because steroids may improve vascular integrity.

Bone marrow transplant has been performed using an HLA-matched sibling, with approximately 70% projected survival. There is a fairly high incidence of one parent matching as well. The most recent approach was described by Gluckman et al. (16): Prenatal diagnosis was performed using chorionic villus cells, and diepoxybutane-induced chromosome breakage was not increased, indicating that the fetus was not a Fanconi's anemia homozygote; HLA typing indicated that the fetus matched the patient. Umbilical cord placental blood was then collected at delivery, cryopreserved, and later used to reconstitute the patient, without having to subject the donor to anesthesia for bone marrow collection. This approach permits use of that material early on, when the donor may be too young to be a bone marrow donor.

In Fanconi's anemia, androgens prolong survival by about four years, although they do not cure the patient (Fig. 10). Some patients gain long-term benefit; I have a patient now who has been taking androgens for eight years, and the 37-year-old man shown in Fig. 7 has been taking androgens for more than 20 years.

Marrow transplant may cure the patient. Standard preparation for marrow transplant for aplastic anemia uses 200 mg/kg cyclophosphamide. When that protocol was used in Fanconi's anemia, there was a high mortality rate, with only a 40% chance of survival, because cyclophosphamide damages DNA, and in these patients DNA of all cells, not just that of their bone marrow, is subject to possible chromosome breakage and repair. Eliane Gluckman in Paris recognized this problem and reduced the cyclophosphamide dose to one-tenth of the normal dose, coupled with 600 Gy of irradiation, and now has an approximately 70% disease-free survival rate in Fanconi's anemia (17). Admittedly, most transplanted Fanconi's anemia patients are not beyond five years yet, and perhaps only a hundred patients with Fanconi's anemia have had marrow transplants worldwide, although marrow transplant may be the cure for aplastic anemia in Fanconi's anemia, I am not sure we yet know quite how to do it right. The outcome in mismatched or unrelated donor transplants is not any better in Fanconi's anemia than in acquired aplastic anemia.

Patients with Fanconi's anemia have other complications. There is approximately a 10% risk of development of leukemia. This may be prevented with marrow transplant. And approximately 10% get other cancers, half of which are liver cancers. The other cancers in the rest of the patients are usually squamous carcinomas, half of which are in women, involving the gynecologic system, and most of the others in the gastrointestinal tract, from tongue to esophagus to stomach to colon to anus. A patient at Mount Sinai Hospital two years ago had esophageal cancer, Fanconi's anemia, and died (18).

How does one treat these patients? The agents used for treatment of their malignancy will interfere with their DNA. Treatment of the leukemia is difficult because normally with myeloid leukemia, which is what these patients have, treatment is for aplasia. Normals recover, but Fanconi's anemia patients may not, because they do not have residual normal stem cells.

Androgens have been implicated as causing the complications. Perhaps they cause some of the liver cancers, but many patients who die of leukemia or other cancers have never taken androgens, and many have never had aplastic anemia. Cancer is the complication of the older patient, and cancers are the things we now warn marrow transplant patients about, because they are not going to have aplasia or leukemia, but they might live longer and get cancer. And the immunosuppressive preparation for transplant might actually exacerbate this tendency.

In 1947 Estren and Dameshek wrote a paper on familial aplastic anemia that was thought not to be Fanconi's anemia. They treated two families with multiple affected siblings who had aplasia (19). More recently it was found that in one of these families there is a bona fide, chromosomally proven case of Fanconi's anemia, and what had been called Estren-Dameshek aplasia is now thought to be Fanconi's anemia without the physical anomalies (20). And, indeed, in an international registry maintained at Rockefeller University, a third of the patients who had aplastic anemia and were homozygotes for Fanconi's anemia, based on chromosome breakage results, had no physical abnormalities. There were also siblings of affected patients who had neither anomalies nor aplastic anemia (21). It appears now that at least a third of Fanconi's anemia homozygotes, and maybe as many as half, do not have physical abnormalities. They have not been diagnosed because the diagnosis was not thought of.

TABLE 5

*Hematopoietic Cultures in Bone Marrow Failure Syndromes*

| Disease | CFU-C | BFU-E | Megakaryocytic | Comments |
|---|---|---|---|---|
| FA | Dec | Dec | — | Decrease even with normal CBC |
| DC | Dec | Dec | — | |
| SD | Dec | Dec | — | |
| KS | Dec* | Nl | — | |
| DB | Nl | Dec | — | Increase with Ep, BPA, or IL-3 |
| Amega | — | Dec | — | |
| TAR | Nl | Nl | Dec | |

Dec = decreased or absent; — = not done; Nl = normal; CBC = complete blood count; Ep = erythropoietin; BPA = burst-promoting activity; IL-3 = interleukin 3.

* CFU-C in KS do not contain mature granulocytes without G-CSF.

See Alter and Young (2) for references.

A note on other inherited disorders: The patient shown in Fig. 11 initially had, at age four, a retroorbital hemorrhage due to thrombocytopenia. Her eye was enucleated. She has a malar rash, a reticular, lacy rash that extends to the

TABLE 6

*Prenatal Diagnosis of Bone Marrow Failure Syndromes*

| Disease | Method | No. affected (studied) |
|---|---|---|
| FA | DEB chromosome breaks, CVS, AF, or FBS | 13 (66) |
| DC | Xq28 RFLP | |
| SD | FBS ANC | |
| KS | FBS ANC | 0 (1) |
| DB | FBS Hb, ADA, BFU-E, Doppler echocardiography | 1 (1) |
| Amega | FBS, platelets | 1 (3) |
| TAR | XRay re: radii | 3 (6) |
| | Ultrasound re: radii | 4 (5) |
| | FBS, platelets | 0 (1) |

ADA = adenosine deaminase, in red cells; AF = amniotic fluid cells; ANC = absolute neutrophil count; BFU-E = burst-forming unit—erythroid; CVS = chorionic villus sample; DEB = diepoxybutane; FBS = fetal blood sample; RFLP = restriction fragment length polymorphism.

See Alter and Young (2) for references.

TABLE 7
*Treatment for Bone Marrow Failure Syndromes*

| Disease | Current therapy | BMT survivors | Possible growth factor |
|---|---|---|---|
| FA | Androgens | 37/84 | GM-CSF?, IL-3?, Ep? |
| DC | Androgens | 0/3 | GM-CSF?, IL-3?, Ep? |
| SD | Androgens, steroids | 0/1 | GM/CSF?, G-CSF? |
| KS | Antibiotics | 1/1 | G-CSF, GM-CSF |
| DB | Steroids | 9/10 | IL-3?, GM-CSF?, Ep? |
| Amega | Androgens | 1/1 | GM-CSF?, IL-3?, Ep? |
| TAR | Platelets | 1/1 | |

Factors with ? will be tested; factors without ? are in use.
See Alter and Young (2) for references.

chest; abnormal fingernails, which an eminent hematologist thought were due to cutaneous candidiasis, but are not; and white spots on her tongue. Is there any nonhematologist who might be able to identify the diagnosis? It is dyskeratosis congenita. It involves all the dermal layers, and about half the patients get marrow failure. They all have dystrophic nails, the reticular pigmentation, and leukoplakia. This was thought to be an X-linked disorder, and indeed there are more than 120 males in the literature. But there are also families in which the condition is clearly an autosomal recessive because there are male and female sibling pairs, and even families with dominant genes not passed through the maternal line in 2 or more generations, as an X-linked disorder would be. These are probably different diseases, different genes with similar manifestations, including a high incidence of aplastic anemia, which is the only reason we hematologists see them (otherwise they are followed by dermatologists). They have a fairly high incidence of cancer and a high mortality rate. The age at which the dermatologic manifestations are first noted is a little hard to ascertain from the literature, but could be childhood to midteens. The aplasia may occur in teenagers, and the cancer in young adults. The projected median survival is about age 35 years. In the Fanconi's anemia group the survival was approximately 16 years. So there is a better prognosis in dyskeratosis congenita, but again, no cure.

## Inherited Marrow Failure Disorders

Some generalities about the inherited marrow failure disorders, most of which (Table 5) I have not discussed, are in order here. In bone marrow hematopoietic cultures from Fanconi's anemia patients, the white cell progenitors, called CFU-C, and the erythroid progenitors (CFU-E and BFU-E) are all decreased. Reports in the literature state that these lower levels hold even in patients whose blood counts are normal. We at the Division of Hematology at Mount Sinai now have an ongoing study of this, and we are hoping to correlate the amount of in vitro growth with counts and with the prognosis, because we would like to be able to tell those who are diagnosed by phenotype and who do not have aplastic anemia whether they are ever going to get it. Dyskeratosis congenita also has a decrease in progenitor cells. We have studied one patient with amegakaryocytosis whose progenitors were decreased.

The thrombocytopenia-absent radii patients have a decrease only in megakaryocyte progenitors. All other lineages are normal, and these are the patients who do improve.

Patients with Shwachman-Diamond syndrome have decreased myeloid progenitors.

Kostmann's syndrome is characterized by normal numbers of CFU-C, but they do not form

TABLE 8
*Inherited Bone Marrow Failure Syndromes*

| Dx | Number | M:F | Aplasia | Leukemia | Cancer | Liver | Prenatal |
|---|---|---|---|---|---|---|---|
| FA | 700 | 1.3 | >90% | 9% | 5% | 5% | Chrom brks |
| DC | 150 | 4.7 | 50% | 0% | 7% | 0% | Xq28 RFLP |
| Amega | 36 | 2.6 | 45% | 6% | 0% | 0% | Plat ct |
| SD | 175 | 1.8 | 20% | 5% | 0% | 0% | ANC |
| DBA | 400 | 1.1 | 0% | 1% | 0.3% | 0% | Hb, ADA, BFU-E |
| KS | 120 | 0.8 | 0% | 2% | 0% | 0% | ANC |
| TAR | 160 | 0.7 | 0% | 0% | 0% | 0% | Plat ct, U/S |

M:F = male/female case ratio; Chrom brks = chromosome breaks; RFLP = restriction fragment length polymorphism; Plat ct = platelet count; ANC = absolute neutrophil count; ADA = red cell adenosine deaminase; BFU-E = burst-forming unit—erythroid; U/S = ultrasound.
See Alter and Young (2) for references.

granulocytes, only monocytes and macrophages. When G-CSF is added to their cultures, they do form granulocytes. When these patients are treated with G-CSF, they make granulocytes in vivo. This finally shows a correlation between laboratory and clinical results.

In Diamond-Blackfan anemia there is an abnormality in the erythroid lineage, but erythroid growth can be stimulated by adding factors euphemistically called burst-promoting activities, known now to be predominantly interleukin 3. It is hoped that soon we will be able to test that agent in these patients.

**Prenatal Diagnosis.** Prenatal diagnosis is possible for most of these syndromes (Table 6). A chorionic villus sample taken at 10 weeks' gestation can be cultured in the presence of diepoxybutane, and chromosome breakage can be identified. In probably 100 or so cases studied at Rockefeller University, about 25% were found to have Fanconi's anemia. For dyskeratosis congenita, a restriction fragment length polymorphism is associated with Xq28 in the X-linked males, but this has not yet been applied prenatally. Amegakaryocytic and thrombocytopenia-absent radii (TAR) might be studied by taking fetal blood and doing a platelet count. In TAR patients, ultrasound would also be useful. Kostmann's syndrome might be diagnosed by finding an absolute neutropenia in utero, but the caution is that at 20 weeks, which is when fetal blood sampling can be done, the normal neutrophil count is about 100/fL, so it takes a very courageous person to identify neutropenia in that context. Diamond-Blackfan anemia has not been tested prenatally, but in one case there was anemia and an abnormality in cardiac stroke volume, in which the marrow turned out to be deficient in red cell precursors. I think that is a kind of tautology, and another cause for the anemia could have been present. Patients with Diamond-Blackfan syndrome, inherited red cell aplasia, often have an elevation in red cell adenosine deaminase, for reasons that are not clear. This assay might be used prenatally, but it has not yet been.

**Treatment.** In therapy for these disorders, androgens should be used for most of the conditions (Table 7). TAR patients can be given platelets; they will improve and not need further treatment. Kostmann's syndrome patients can be given G-CSF. We hope soon to test interleukin 3 in patients with Diamond-Blackfan syndrome and perhaps some of the other syndromes.

To reiterate (Table 8), Fanconi's anemia patients may be physically normal, probably comprise at least 5% of all patients with "acquired aplasia," and have a preleukemic and premalignant propensity. Patients with dyskeratosis congenita probably are underdiagnosed because in those with initial aplasia, dyskeratosis was not diagnosed until they were sent to a dermatologist or the diagnosis was thought of by hematologists. No leukemia has been reported, but aplasia and malignancies have been. Amegakaryocytic patients comprise a small number of cases with a high incidence of aplasia and leukemia. Patients with Shwachman-Diamond syndrome may develop aplasia or leukemia. Those with Diamond-Blackfan syndrome develop leukemia but no aplasia, as do Kostmann's syndrome patients. TAR patients have none of these complications.

Those of us who try to study these disorders hope that we can learn something from the single cytopenias that will help us to understand aplastic anemia, and we hope that this concept of thinking of a genetic background for an acquired aplasia can work both ways: The acquired anemias may have a genetic background; in the clearly inherited anemias, aplasia may actually be stimulated by some of the drugs, chemicals, viruses, and other exposures that cause acquired aplastic anemia.

## References

1. Alter BP. The bone marrow failure syndromes. In: Nathan DG, Oski FA, eds. Hematology of infancy and childhood. 3rd ed. Philadelphia: WB Saunders, 1987; 159–241.
2. Alter BP, Young NS. The bone marrow failure syndromes. In: Nathan DG, Oski FA, eds. Hematology of infancy and childhood, 4th ed. Philadelphia: WB Saunders, 1992.
3. Sleijfer DT, Mulder NH, Niewig HO, Anders GJPA, Gouw W-L. Acquired pancytopenia in relatives of patients with aplastic anaemia. Acta Med Scand 1980; 207:397–402.
4. Camitta BM, Thomas ED, Nathan DG, et al. Severe aplastic anemia: a prospective study of the effect of early marrow transplantation on acute mortality. Blood 1976; 48:63–70.
5. Nathan DG. Introduction: hematologic diseases. In: Wyngarden JB, Smith, JH, eds. Cecil textbook of medicine, Vol I. Philadelphia: WB Saunders, 1985:866–870.
6. Spangrude GJ, Heimfeld S, Weissman IL. Purification and characterization of mouse hematopoietic stem cells. Science 1988; 241:58–62.
7. Kurtzman G, Frickhofen N, Kimball J, Jenkins DW, Nienhuis AW, Young NS. Pure red-cell aplasia of 10 years' duration due to persistent parvovirus B19 infection and its cure with immunoglobulin therapy. N Engl J Med 1989; 321:519–523.
8. Yunis AA. Mechanisms underlying marrow toxicity from chloramphenicol and thiamphenicol. In: Silber R, LoBue J, Gordon AS, eds. The year in hematology. New York: Plenum Press, 1978:143–170.
9. Li FP, Alter BP, Nathan DG. The mortality of acquired aplastic anemia in children. Blood 1972; 40:153–162.

10. Camitta BM, Storb R, Thomas ED. Aplastic anemia: II. Pathogenesis, diagnosis, treatment, and prognosis. N Engl J Med 1982; 306:712–718.
11. Deeg HJ, Self S, Storb R, et al. Decreased incidence of marrow graft rejection in patients with severe aplastic anemia: changing impact of risk factors. Blood 1986; 68:1363–1368.

11a. Anasetti C, Doney KC, Storb R, et al. Marrow transplantation for severe aplastic anemia: long-term outcome in fifty "untransfused" patients. Ann Intern Med 1986; 104:461–466.

12. Gordon-Smith EC. Treatment of severe aplastic anemia by bone marrow transplantation. Hematol Oncol 1987; 5:255–263.
13. Shahidi NT, Gerald PS, Diamond LK. Alkali-resistant hemoglobin in aplastic anemia of both acquired and congenital types. N Engl J Med 1962; 266:117–120.
14. Auerbach AD, Adler B, Chaganti RSK. Prenatal and postnatal diagnosis and carrier detection of Fanconi anemia by a cytogenetic method. Pediatrics 1981; 67:128–135.
15. Auerbach AD, Alter BP. Prenatal and postnatal diagnosis of aplastic anemia. In: Alter BP, ed. Methods in hematology: perinatal hematology. Edinburgh: Churchill Livingstone, 1989:225–251.
16. Gluckman E, Broxmeyer HE, Auerbach AD, et al. Hematopoietic reconstitution in a patient with Fanconi's anemia by means of umbilical-cord blood from an HLA-identical sibling. N Engl J Med 1989; 321:1174–1178.
17. Gluckman E. Bone marrow transplantation for Fanconi's anaemia. Bailliere's Clin Haematol 1989; 2:153–162.
18. Gendal ES, Mendelson DS, Janus CL, Schlossberg I, Vogel JM. Squamous cell carcinoma of the esophagus in Fanconi's anemia. Dysphagia 1988; 2:178–179.
19. Estren S, Dameshek W. Familial hypoplastic anemia of childhood. Report of eight cases in two families with beneficial effect of splenectomy in one case. Am J Dis Child 1947; 73:671–687.
20. Li FP, Potter NU. Classical Fanconi anemia in a family with hypoplastic anemia. J Pediatr 1978; 92:943–944.
21. Auerbach AD, Rogatko A, Schroeder-Kurth TM. International Fanconi Anemia Registry: relation of clinical symptoms to diepoxybutane sensitivity. Blood 1989; 73:391–396.

# *The Future Of Medicine*

It is not often that I have had the opportunity to read essays written by high school students other than those of my own children. From time to time, I have been asked to review the scientific writings of students who have submitted reports on their research undertakings. These are unusual students, highly motivated and over-achievers. Some may be driven by their parents to excel in these efforts, but many accept the challenges of their academic and scholarly work with high regard for the sense of fulfillment and satisfaction they themselves experience in discovering their own talents and maturation. It therefore occurred to me that many of my professional colleagues might be interested in sharing with me the experience and joy which comes from being a teacher to even one student who has completed a task that clearly is noteworthy.

I did not seek out the essay written by Andy Schanzer herein published in the Journal. His father, Dr. Harry Schanzer, is a distinguished member of the Department of Surgery of the Mount Sinai School of Medicine. It is clear that this parent is a proud one indeed and rightfully so, as the reader of this essay will conclude. Dr. Schanzer knew of my interest and that of the Journal in promoting the publication of scholarly writings by students in general. With this in mind he asked me if I would be willing to read an essay recently written by his son for his English class. The outcome of this exchange is for all to see.

As an individual with a long-time interest in education, and an active participant in many phases of the process itself, I was especially impressed by Andy Schanzer's sensitivity and enthusiasm. Perhaps one might have expected such expression from the son of a surgeon, but I do not believe that for a moment. Moreover, I was very much impressed by the comments made by the teacher, Michael W. DiGennaro. These comments carry very important implications with respect to the future of medical education and what it will take to attract those who might aspire to the practice of medicine. And to that end, we must recognize the contributions of those teachers of our youth to whom we have given this awesome responsibility. For too many and for too long this recognition has not been on our educational agenda. Surely, our personal and professional instincts suggest that we pay homage to the teacher and student for a job well done.—*Sherman Kupfer*

# Reflections on Richard Selzer's *Mortal Lessons: Notes on the Art of Surgery* by An Adolescent Aspiring Surgeon

ANDRES SCHANZER

**The hidden geography of the body is a Medusa's head one glimpse of which would render blind the presumptuous eye.—Selzer, p. 24.**

1. AFTER reading the first chapter of Richard Selzer's *Mortal lessons: notes on the art of surgery* (New York: Simon & Schuster, 1987) I constantly found myself rereading this particular passage, not exactly sure of what I was looking for. I believe that Selzer is saying that, by opening the body, one is in turn violating nature. I, interested in being a doctor all of my life, have also often thought about this question—not so much in Selzer's sense of violation, but more so concerning the question of futility. What I mean is that sometimes I wonder if being a surgeon makes one blind to the fact that many times a person is going to die regardless of what he or she does; this is not the doctor's fault, nor anyone else's. It is simply nature. Furthermore, I also wonder whether a doctor really does anything at all. Is someone "supposed" to live or does a doctor "save" him? Did the doctor make a mistake or simply had the patient's "time come?" It is questions such as these that in one sense make me re-ask myself "is medicine for me?" but, on the other hand, intrigue me, make me hunger for the scalpel. I do not intend to change nature's course, nor to play God. My only desire is, if humanly possible, to enter a profession where my skill will be beneficial to those in need and where I will be able to do my utmost to make a difference. If I decide that medicine is indeed this special profession that I have just described, then do not be surprised to find me, ten years from now, glimpsing the hidden geography of the body Richard Selzer speaks of.

2. During one of our most insightful class discussions, one of my peers stated something to the effect that "doctors seem to be very insensitive; they don't even care if their patients die." I responded, "Don't you understand that when you are in a profession such as medicine, in order to survive, you must gain a degree of callousness to death?" After returning home I began to question myself. I know that regardless of how good a doctor is, if he is in a difficult field, inevitably he will lose patients. I also know that it is crazy to cry and mourn each lost patient when you're doing your best to save lives. However, after reevaluating my original statement, "a doctor must have a tolerance for death," I began to wonder if anyone really should ever be callous about something as serious as death. A person's passing away is probably one of the major events that another human must face. Is it right for someone to see a "flat line" on a heartbeat monitor and feel nothing besides a possible slight disappointment with himself? I am not sure.

It is common to hear doctors described as "wounded healers." I believe that this title is quite appropriate. For to be a doctor, one must unfortunately lose some elements that go hand in hand with being a human being. I think it is a very false, ignorant view for someone, such as my classmate, to see doctors as insensitive. One must imagine how difficult it is to take on the heavy burden of being a doctor. I, along with Richard Selzer, see medicine as the noblest of all professions. In summation, I say, "To all doctors, you have my respect!"

---

The author is a student at Mamaroneck High School, Mamaroneck, NY. Address reprint requests to the author at Mount Sinai Medical Center, Box 1259, Department of Surgery, One Gustave L. Levy Place, New York, NY 10029.

3. Allow me to stray from the essence and depth of the book and look at it in a purely analytical sense. It would be much more difficult to use the entire book to illustrate several of Richard Selzer's writing techniques, so I will stick to only three pages.

A bit of background: This short story is about a man recently returned from Guatemala, from the excavation of Mayan ruins. In his left upper arm Selzer finds an organism of sorts, a creature.

The description is beautifully intricate. Although somewhat frightening, it is as though the reader is an onlooker at Selzer's side. For example, he describes his first view: "His left upper arm wears a gauze dressing which, when removed, reveals a clean punched-out hole the size of a dime, the tissues about the opening are swollen and tense. A thin brownish fluid lips the edge, and now and then a lazy drop of the overflow spills down the arm. An abscess, inadequately drained . . ." (p. 20). What a picture!

Selzer then goes on to describe, with his usual clever and somewhat humorous tone, what he sees after enlarging the opening. For instance, to show the extent of his surprise, he uses this line: "What happens next is enough to lay Francis Drake avomit in his cabin. No explorer ever stared in wilder surmise than I into that crater from which there now emerges a narrow gray head whose sole distinguishing feature is a pair of black pincers" (p. 20).

Selzer's reference to Sir Francis Drake is both creative and typical of his unique style. Following this excerpt, Selzer further describes this "Mayan Devil" (p. 20) and the entire scenario without leaving out any minute detail. He then goes on to add his own insights and views. He thinks to himself, "[It] would soon burst free to fly about the room, with horrid blanket-wings and iridescent scales, raking, pinching, injecting God knows what acid juice" (p. 20). He tries to show the reader that doctors too have fears.

Throughout the book, Selzer's asides show how highly he esteems the medical profession. For example, "with all the ritual deliberation of a High Priest I advance . . ." (p. 20). Here is surgeon as savior indeed.

Another technique Selzer often uses is comparing the body to nature. For instance, he chooses the metaphor of a lake to describe the horrifying abscess that his patient suffers from. Upon seeing the creature in his patient's arm, Selzer says, "A faint disturbance in the lake, and once again the thing rises" (p. 21).

Selzer also describes the feelings that run though him while working at the operating table. Upon capturing the organism, he writes, "Pinned and wriggling, he is mine. . . . A hatred seizes me, but such a detestation as would make of Iago a drooling sucktit" (p. 21) (again the same type of humor as the reference to Francis Drake). "It is the demented hatred of the victor for the vanquished . . ." (p. 21).

I hope that by this analytical look at Selzer's style one can gain more of an idea of exactly what is behind his writing.

4. "He has found the spot, and for the next half hour he remains thus suspended above the patient . . . holding the pulse of the woman beneath his fingers, cradling her hand in his . . . from the foot of the bed, where I stand, it is as though he and the patient have entered a special place of isolation, of apartness, about which a vacancy hovers, and across which no violation is possible. . . . All at once I am envious—not of him . . . but of her. I want to be held like that" (p. 34).

This part of the book marks a significant event. It is the first time that Richard Selzer describes his personal, intimate feelings in regard to something other than the wonders of the body. Selzer is showing the reader that even he, who definitely has a love for medicine, gets fed up at times with playing the role of healer. He would like the chance to sit back for once and be cared for.

He appreciates his profession so much that he, instead of being the one who administers the help, would like to be the person to benefit. He yearns for the security he imagines one feels in the hands of a skilled doctor. This does not make him a pompous person, nor does it mean that he is unhappy with his profession. I think there is a very gratifying factor about being a doctor which he recognizes. After all, you're saving lives. Still, one cannot overlook the negatives—exhaustion, stress, pressure, etc. Most doctors are well aware of both sides of their profession. I cannot even begin to count the number of times my father, a surgeon, has come home, sat in a chair, closed his eyes, and said, "Son, I know I'm doing a good thing, yet sometimes I just wish it was someone else standing over me in deep concentration while I am deep asleep. The medical profession is a good one but be sure it is the right one for you." Dad, I tell you honestly that I really do think it is, and do not think I don't understand or sympathize with you and our mutual friend, Richard Selzer. It's simply that in this profession, if you are good and work hard, there is no question that the ups outweigh the downs.

5. "But even now, years later, I start from my bed as I hear the hissing of those bones, and it

does not matter what you say, or if you think that what I've told is true. It matters that I have been changed by it. That I am not the same as I was" (p. 58). The medical profession can do this to you more so than any other type of employment. You can see, hear, or feel one thing and be changed for the rest of your life. To see a person struggling, fighting one on one with death, drenched in sweat for several hours, can alter your entire view of both life and death, yet our doctors of today are faced with these situations every day and still people refer to them as insensitive professionals, hungry for money. Someone who devotes eight years studying to save lives, I believe, gives a whole new definition to sensitivity. One must understand that as a person learns more and more about the human body, they begin to see it as a wonder, an intricate machine. Clotting is no longer a scab on a cut, it is a process. Fibrinogen attacks the sore, fibrin strands form across the wound releasing thromboplastin. The thromboplastin forms a network of thrombin and solidifies. Finally, the end-product is formed, a scab. And this many people think is only a hard little purple thing. A doctor knows the secrets. He or she values life, holds it sacred.

I witnessed a kidney transplant five months ago. I was right there, scrubbed, and looking right over my father's tense shoulders. It was beautiful! He was my hero—slicing, clamping, tearing, sewing—all with such precision. From just that three-hour exposure, I believe that I was changed. I'm not sure if you would call it a maturation, but it was more of an appreciation for life, for my father, and for doctors as a whole.

This was my first exposure to surgery and I already felt the impact of manipulating the human body. It was exhilarating. And, whether you believe me or not, I truly believe, and that I believe of course is what matters, that "I am not the same as I was" (p. 58).

6. "The parts of the urinary tract, namely the kidneys, ureters, and bladder—are a tracery of tissues and ducts of such fineness as would lay mad a master plumber, more, a Venetian glassblower" (p. 28). Selzer has such a love and respect for the body that one almost feels exhilarated while reading, exhilarated in the sense that the writer makes the reader appreciate and value the complexity of what he or she actually is.

No car or airplane or refrigerator that I know of has ever run for 40 years, let alone 90. Yes, we all live in the computer age. We can split atoms that we cannot see or feel, but still, we cannot create an efficient, long-term mechanism that can consistently carry out simple functions that we humans do every day.

Alright, fine, we cannot create an entire machine, so let us focus simply on one part of this unreproducible mechanism, for example, the skin. We've all got it. Quite a lot of it, in fact. Yet, after living on this earth for thousands and thousands of years, not one human has been able to synthesize the common protective surface we refer to as skin. We can take it from another part of the body. We can culture it so it grows. But what we cannot do is produce it.

So what does all this tell us? It humbles us. It says to men and women something as simple as this: You are doing well, but you are not doing great. The human body is a wonder. I accept this, Richard Selzer does, but still, many do not. No one understands it all, for the body is too brilliant.

As Selzer describes the urinary tract, he says "it puts the best of plumbers to shame" (p. 78). I agree. And I will further venture to say that the body will continue to outdo the smartest in any profession. I say this because I, like Selzer, have a love and hold a very high respect for my "self."

7. "Imagine God as a tailor. His shelves are lined with rolls of skin, each with its subtleties of texture and hue. . . . Now engage the irreverence to consider that, either out of the tedium born of infinity, or out of mere sly parsimony, he uses for the occasional handicraft a remnant of yard goods, the last of an otherwise perfect bolt, dusty, soiled, perhaps a bit too small or large, one whose wool is warped or that is cut on the bias. I have received many such people in examination rooms. Like imperfect postage stamps, they are the collector's items of the human race" (p. 113).

I think that Selzer is saying very much in this paragraph. In fact, I find it to be somewhat philosophical. He speaks of God creating babies and sealing their bodies in skin which is simply rolled up upon his shelves. There are rolls of different colors and textures. Arbitrarily the organs of a baby's body are encased in whichever roll God should choose. Selzer then goes on to state that occasionally God uses the end of a roll, a torn piece possibly, a wrinkled piece, or simply a piece either too big or too small. The people (for lack of a better word) wearing these pieces are the irregulars, the humans who waiver drastically from normality—the ugly, the horrid, the wretched. He looks at these conditions from a scientific point of view and in turn appreciates them. He compares them to imperfect postage stamps, the ones with the most value, the most important. These, he says are "the collector's items of the human race" (p. 113).

Selzer is both a sensitive and an observant person. He feels that everyone, regardless of appearance, belongs to and holds a place in society. After all, scientifically speaking, a 5′5″ girl with blond hair and blue eyes is much less special and unique than a bald dwarf with one hazel eye and the other permanently shut since birth.

I won't belabor this topic any further. The moral of the story is simply this: before looking down at someone as a result of his or her appearance, take a moment to think, and if a moment is not long enough, then think long and hard. For every person is beautiful in one way or another, whether they match society's standards or not. Now that you have learned your lesson, be careful, because the society of this day and age is way too cruel. Don't let yourself be sucked in!

8. The chapter in Selzer's book titled "The Corpse" is a serious one dealing with quite a serious topic, death. I do not think about death much; nor do I fear it. What it boils down to is that one day my time is going to come. I can honestly say that I hope this time does not come for a long while, but when it does, hey! that's it, I guess.

In his story of the corpse, Selzer raises an interesting question that I was forced to think about. What should be done with the body? Cremation, burial, embalming? I personally came to this conclusion: If there is something out there past death, then I can tell you one thing, I sure as hell don't think I'm going to need the same old body there—this means that to me it's really unimportant either way. If possible, I'd give my organs to science—I'm sure they would be of more use to scientists than to me.

If for some reason none of my organs would benefit the scientific world, then I leave "me" to my family. Do what you feel most comfortable with. Just remember, my spirit is somewhere far off and that is what I feel matters, not the dead skin, not the still heart, and not the rigor-mortised muscles either. I hope I don't sound negative. Everyone's entitled to their own beliefs. If some think that their bodies should be preserved as best as possible, then I can respect that. For me though, I still believe, and maybe this will be subject to change, that a man or woman's spirit always lives on forever independently of his or her remains. So remember, regardless of what you do with "me," I'll always be around.

9. I must confess that Selzer's chapter on smoking was very disappointing. In fact, I think it caused me to lose a little respect for this man I had grown to admire. Throughout the entire book he talks about how wonderful the "sacred body" is. He understands it so well, both scientifically and as a person. Yet he tells the reader, in this chapter, that he smokes. What makes his addiction worse than that of others is the fact that he is well aware of the damage to his body that he is self-inducing.

Selzer claims that the reason he smokes is to remind himself of his own existence. He even goes to the extent of saying, "To deny me my smoke is to extinguish me as utterly as would death itself" (p. 150). Allow me to be perfectly candid. I think that this excuse is totally ridiculous. Come on, Dr. Selzer, there are plenty of ways to remind yourself of existence. Ski down a mountain full speed naked of goggles, go jogging on a beach all alone at six in the morning during one of those hot summers, feel the sand in between your toes, work up a sweat, and throw your body into the waves. This I tell you is life, not "drawing deeply upon a cigarette to fill the tracheobronchial tree with smoke" (p. 150).

Well my friend Selzer, I can only tell you this. I'm very disappointed. For a man of your stature, who holds the human body in such high regard, it is very disturbing to hear that of all people, it is you who voluntarily damages it.

10. "I see something! . . . I see a movement—a small one. But I have seen it. . . . I see that it is the hub of the needle in the woman's belly that has jerked. . . . It is the fetus struggling against the needle, struggling? . . . That cannot be. . . . It is merely reflex. I point to the needle. It is a reflex, says the doctor" (p. 158).

The entire chapter dealing with abortion was quite frightening. This excerpt especially made me think. Let me begin by saying I am prochoice. But the fact that the fetus as young in age as 24 weeks has reflexes raises questions even in my prochoice mind. Murder of any human being is wrong. So therefore, I guess the question really is, when is the fine line of becoming a human being actually crossed? If a group of human cells have only reflexes, does this in turn mean it is a human being? I tend to think not, for it has been established that the fetus "feels no pain, cannot feel fear, and has no motivation" (p. 158). I tell myself over and over, the fetus feels no pain, therefore, abortion is okay. Then why do I tremble when Selzer refers to the five others in the room and says, "we are not six, I think. We are seven" (p. 158). The reason is obvious. I am not sure. How can I be? Maybe abortion is wrong. In a sense denying a life is frighteningly similar to ending a life. One must consider that last statement carefully. It is quite serious; so I leave these questions to you: Which is worse? What is right and wrong? You decide.

**11.** In retrospect, I've learned many things about both the art of surgery and myself. I've also learned to love and appreciate the complex machinery that comprises "me." No longer will I look at the body as a normal everyday thing that is part of nature. I've realized that it is intricate, fascinating, and valuable.

Reading Selzer's book has strengthened my desire to aspire and become a part of the medical profession. This desire has transformed from what began as a childhood dream to a realistic goal. I want to save lives and do what I can to help the society in which I have grown, my generation. Medicine, I feel, is the best way to achieve this. As long as I have my mind set, I am confident that I will be able to do great things. I *will* make a difference.

Finally, I would like to thank Richard Selzer for helping me to learn what we are. My journey through his creation, *Mortal Lessons,* has been insightful and exciting. I have realized that Selzer's assertion right in the beginning of the book is indeed true: "The surgeon is a priest."

---

## *Comment*

ANDY Schanzer's reactions to Richard Selzer's book were always interesting to read because Andy gives the reader many glimpses into his own sensibility. What makes this particularly fascinating is his own disposition toward medicine. He's the only one in the class to express such an interest. When I first started teaching over 20 years ago, it seemed that at least half of my brightest students aspired toward a career in medicine. In fact, I began using essays by Richard Selzer primarily because so many young people wanted to be doctors and I wanted to show them that scientific thinking and imaginative writing did not have to be mutually exclusive.

Over the years I've noticed that fewer and fewer of the advanced, high-achieving students wanted to become doctors. I don't know exactly why this decline has occurred or if such aspirations at the secondary level correspond at all to the numbers of students who eventually enter the profession. From what I have heard, it's still extremely difficult to get into medical school. It just seems that the dream at an early age of becoming a doctor is not what it used to be. I can only assume that the rigors of getting into medical school, the intellectual and psychological demands of studying medicine, and the extraordinary personal sacrifices that go into establishing and maintaining a practice have somehow made an impression on high-school students and have offset the more romantic, idealistic ideas that were once associated with being a doctor. Students today can conceive of quicker, easier roads to success and personal fulfillment.

My inclusion of Selzer in the study of English has little, if anything, to do with changing attitudes toward the medical profession. I simply want students to be exposed to the writing of a man who can think and perform as a doctor, yet use language to express such thoughts as a poet would. Although my emphasis has been literary, the discussion invariably gets into controversial issues such as Andy discusses, particularly the nature of the relationship between a doctor and his or her patient.

That Andy is willing to pursue the dream of becoming a doctor and accept all of its challenges is, indeed, admirable. The *Journal*'s interest in publication of the essay indicates its sensitivity to Selzer's complex sensibility, one that can harmonize the multiple roles a physician performs—surgeon, priest, and artist.—*Michael W. DiGennaro, Mamaroneck High School, Mamaroneck, NY.*

# List of *Journal* Reviewers in 1991

*Each year the JOURNAL relies on its reviewers as the lynchpin in offering its readers worthwhile, timely, and interesting peer-reviewed articles. The JOURNAL staff thanks the following reviewers for their time and intellectual effort on manuscripts submitted to the publication in 1991.*

Bernard Adelsberg
Alan M. Aron
Peter Banks
Joseph Eisenman
Richard A. Frieden
Steven M. Fruchtman
J. Lester Gabrilove
Harriet S. Gilbert
Paul L. Gilbert
James Godbold
Ezra M. Greenspan
Adrian Greenstein
Richard S. Haber
Noam Harpaz
Shalom Z. Hirschman
Barry W. Jaffin
Thomas A. Kramer
Isadore Kreel
Leslie Kuhn
Sherman Kupfer
Richard Lasser
Kenneth V. Lieberman
Charles J. Lockwood
Leonard I. Malis
John Mandeli
Steven B. Markowitz
Zvi M. Marom
Myron Miller
Daniel A. Moros
Daniel H. Present
Edward L. Raab
Jack G. Rabinowitz
Allan P. Reed
Rosamond Rhodes
Sheila H. Roman
Allan E. Rubenstein
Jack Rudick
David B. Sachar
Henry Sacks
Martin H. Savitz
Clyde Schechter
Donald A. Smith
Harry Spiera
Barry D. Stimmel
Alvin S. Teirstein
Allen Unger
Melvin H. Van Woert
Samuel Waxman
Jerome D. Waye
J. Lawrence Werther
Michael F. Wesson
Stuart H. Young
Barnett Zumoff

# Mount Sinai School of Medicine Faculty Awards and Prizes

Alexander Richman Commemorative Award for Humanism in Medicine, 1991
**Kurt W. Deuschle, MD**

Alice Hamilton Award, American Public Health Association, 1990
**Irving J. Selikoff, MD**

American Association for the Advancement of Science, Fellow, 1991
**Kurt W. Deuschle, MD**

De Roaldes Medal,
American Laryngological Association
**Joseph Goldman, MD**

Duncan Clark Award, Association of Teachers of Preventive Medicine, 1990
**Kurt W. Deuschle, MD**

E. Mead Johnson Award for Research in Pediatrics, Society for Pediatric Research
**Gregory Grabowski, MD**

E. Mead Johnson Young Investigator Award, Society for Adolescent Medicine
**Angela Diaz, MD**

E. Mead Johnson Young Investigator Award, Society for Adolescent Medicine
**Doris Pastore, MD**

Heinz Hartman Award for Contributions to Psychoanalysis, 1991
New York Psychoanalytic Society
**Martin H. Stein, MD**

Henry Christian Award for Excellence in Research, American Federation for Clinical Research
**Karen Segal, PhD**

Interboro Leukemia Organization Award, 1991
**Vesna Najfeld, PhD**

International Award for Cancer Prevention, Italian Cancer Society, 1990
**Irving J. Selikoff, MD**

Master of Gastroenterology, American College of Gastroenterology, 1991
**Jerome D. Waye, MD**

Mount Sinai Endocrinologist of the Year Award, 1991
**Peter N. Graves, PhD**

Outstanding Devotion to Teaching Award, Department of Medicine, Mount Sinai School of Medicine, 1991
**Jerome D. Waye, MD**

René Dubos Award, 1990
**Irving J. Selikoff, MD**

Sebetia-Ter National Award (Italy), 1990
**Irving J. Selikoff, MD**

Solomon Berson Teaching Award, 1991
**Louis M. Aledort, MD**

University Faculty Senate of the City University of New York, Resolution of Appreciation, 1991
**Charlotte F. Muller, MD**

Young Turks Award, American Society for Clinical Investigation, 1991
**Lloyd Mayer, MD**

# Subject, Author, and Title Index to Volume 58

Abstracts: alcoholism, 58(4):347; cardiology, 58(4):347–350; clinical immunology, 58(4):350–351; endocrinology, 58(4): 351–355; gastroenterology, 58(4):355–356; geriatrics, 58(4):356–358; hematology, 58(4):358–359; hypertension, 58(4):359; infectious diseases, 58(4):359–360; liver diseases, 58(4):360–361; molecular medicine, 58(4):362; by Mount Sinai School of Medicine students, 1991, 58(4):365–366; from Mount Sinai School of Medicine, Dept of Medicine, 1991, 58(4):347–365; neoplastic diseases, 58(4):362–363; oncology, medical, 58(4):361; pulmonary diseases, 58(4):363–364; renal diseases, 58(4):364–365; rheumatology, 58(4):365; spinal cord damage, 58(4):365
Acquired brain lesions and psychiatric illness: possible role of dopaminergic systems, TA Kramer, A Merriam, P Harvey, MD Lisoskie, 58(4):324–327
Acute herpetic neuralgia and postherpetic neuralgia: clinical features and management, BS Galer, RK Portenoy, 58(3):257–266
Affective disorders, and seizures, 58(4):310–323
AIDS and surrogate decision making, DA Hansell, 58(5):375–378
AIDS: see HIV
Aisen PS: Elderly psychiatric patient, 58(1):85–90
Alexander Richman lecture: Medical ethics, a broader definition, SM Wolfe, 58(5):455–461
Alcoholism: abstracts, 58(4):347
Alter B: Bone marrow failure disorders, 58(6):521–534
An adolescent aspiring surgeon's reflections on *Mortal Lessons: Notes on the Art of Surgery* by Richard Selzer, with comment by Michael W. DiGennaro, A Schanzer, 58(6):536–540
Analgesics and pain: 58(2):101–108; 58(3):221–228
Anesthesia: what the internist needs to know, AB Leibowitz, JH Silverstein, 58(1):9–18
Antidepressants and pain: 58(3):221–228
Antimicrobial therapy, E Neibart, 58(1):19–27
Anvaripour PL, Jacobson L, Schweiger J, Weissman GK: Physician-nurse collegiality in the medical school curriculum: exploratory workshop and student questionnaire, 58(1): 91–94
Approach to the heterogeneity of diabetes in adults, SH Roman, 58(4):280–286
Art and science of consultation in medicine, PL Gilbert, ed, 58(1):1–94
Arthritis of the lumbar spine, GV Gordon, 58(2):109–114
Arthroscopic microdiskectomy, P Kambin, 58(2):159–164
Asthma: and atopy, 58(6):494–499; in children, 58(6):500–506; differential diagnosis, 58(6):466–471; mast cells in, 58(6): 472–482; occupational, 483–493; pulmonary function testing, 58(6):507–520
Asthma in children and its management: an update, D Lowenthal, M Kattan, 58(6):500–506
Asthma: new concepts, ZM Marom, ed, 58(6):463–520
Asthma room, the, AS Teirstein, 58(6):463–465
Atopy: in asthma; 58(6):494–499
Automated percutaneous lumbar diskectomy, G Onik, 58(2): 151–158
Awards: *Mount Sinai Journal of Medicine,* 1990, 58(3):iii; Mount Sinai School of Medicine, 1990/91, 58(3):ii; Mount Sinai School of Medicine Faculty, 58(6):542

Barbasch A, moderator, R Gorlin, D Callahan, B Baumrin, discussants: Doctor's dilemmas: bioethical challenges, panel discussion, 58(5):443–451
Baumrin B, moderator: Drug use: public policy, panel discussion, 58(5):441–442
Baumrin B: see Barbasch A
Bone marrow failure disorders, B Alter, 58(6):521–534
Bonica JJ: History of pain concepts and pain therapies, 58(3):191–202
Brock DW: Surrogate decision making for incompetent adults: an ethical framework, 58(5):388–392
Brown LK: Pulmonary function testing in bronchial asthma: standard and emerging techniques, 58(6):507–520
Buchbinder D: see Kaplan AS
Burton CV: Causes of failure of surgery on the lumbar spine: ten-year follow-up, 58(2):183–187

Calandri C: see Rand JH
Callahan D: see Barbasch A
Cardiac disease: 58(1):48–57; abstracts: 58(4):347–350
Causes of failure of surgery on the lumbar spine: ten-year follow-up, CV Burton, 58(2):183–187
Cells, mast, and asthma: 58(6):472–482
Changing controversy over surgical resection margins for stage I cutaneous melanoma, MS Green, 58(4):341–346
Cherry SH: Pregnant patient: need for surgery unrelated to pregnancy, 58(1):81–84
Children: and asthma, 58(6):500–506; low back pain, 58(2): 115–120; pain management, postoperative, 58(3):247–256; surrogacy for, 58(5):382–383, 393–397; see Infants and children
Clinical immunology: abstracts, 58(4):350–351
Cocaine use in New York State; 58(5):406–411
Conn M: Patient with liver disease, preoperative evaluation, 58(1):75–80
Conservative management of low back pain, RA Frieden, 58(2):97–100
Consultation by internists: 58(1):1–94

Des Jarlais DC: see Frank B
Diabetes mellitus: strategies for the 1990s, SH Roman, ed, 58(4):273–366
Diagnosis and treatment of myofascial pain, H Kraus, AA Fischer, 58(3):235–239
Difference between acute and chronic pain, KP Grichnik, FM Ferrante, 58(3):217–220
Differential diagnosis of asthma, AS Teirstein, 58(6):466–471
DiGregorio GJ: see DiPalma JR
DiPalma JR, DiGregorio GJ: Management of low back pain by analgesics and adjuvant drugs, 58(2):101–108
Diskectomy: 58(2):147–168
Doctors' dilemmas: bioethical challenges, panel discussion, A

Barbasch, moderator, R Gorlin, D Callahan, B Baumrin, discussants, 58(5):443–451
Dreiling DA, death: (58)6: inside back cover
Drug policy in New York City, SC Joseph, 58(5):421–422
Drugs and the social body, PS Mann, 58(5):427–432
Drugs for pain: 58(2):101–108; 58(3):221–228
Drug use and public policy, JB Murtaugh, 58(5):423–426
Drug use: license vs. prohibition, 58(5):403–451; cocaine epidemiology, 406–411; history of, 403–405; paternalism and, 412–415; pharmacology, 416–420; and public policy, 421–436
Drug use: public policy, panel discussion, B Baumrin, moderator, 58(5):441–442
Drug use: social and scientific background, panel discussion, DA Moros, moderator, 58(5):437–440

Editorials: Art and science of consultation in medicine, S Kupfer, 58(1):1; Lumbar disk disease: controversies in neurosurgery, MH Savitz, 58(2):95–96; The future of medicine, S Kupfer, 58(6):535
Eisenbarth G: Prediction and prevention of type I diabetes, 58(4):274–279
Elderly patients: and diabetes, 58(4):287–292; and intervention, 58(4):328–340; surrogacy for, DE Meier, 58(5):385–387; with psychiatric disorders: 58(1):85–90; see Geriatrics
Elderly psychiatric patient, PS Aisen, 58(1):85–90
Embolism: 58(1):28–36
Endocrine disease: 58(1):58–68; abstracts, 58(4):351–355
Epidemiology of cocaine use in New York State, B Frank, DC Des Jarlais, R Marel, J Schmeidler, M Maranda, 58(5): 406–411
Ethics of surrogate decision making, panel discussion, DA Moros, R Rhodes, moderators, 58(5):398–402
Ethics: 58(5):367–461; bioethical challenges, 443–451; and drug use, 403–442; personal ethics, 452–454; social ethics, 455–461; and surrogacy, 369–402

Faculty awards and prizes, Mount Sinai School of Medicine, 58(6):542
Fager CA: Role of laminectomy in lumbar disk herniation, 58(2):133–138
Fahy GM: Short-term and long-term possibilities for interventive gerontology, 58(4):328–340
Ferrante FM: see Grichnik KP
Findling RL: see Post RM
Fischer AA: see Kraus H
Foster parents as surrogates for infants and young children, JD Moreno, 58(5):393–397
Fox DM: History of drug use and drug control, 58(5):403–405
Frank B, Des Jarlais DC, Marel R, Schmeidler J, Maranda M: Epidemiology of cocaine use in New York State, 58(5): 406–411
Frieden RA: Conservative management of low back pain, 58(2):97–100

Galer BS, Portenoy RK: Acute herpetic neuralgia and postherpetic neuralgia: clinical features and management, 58(3):257–266
Gastroenterology, abstracts, 58(4):355–356
Geriatrics, abstracts, 58(4):356–358; see Elderly patients
Gilbert PL: ed, Art and science of consultation in medicine, 58(1):1–94; Introduction: In omnia paratus–ready for all things, 58(1):2; Internist as preoperative consultant: risk assessment and management, 58(1):3–8; Patient with endocrine disease, preoperative evaluation, 58(1):58–68
Gilbert PL, Stein R: Patient with chronic renal disease, preoperative evaluation, 58(1):69–74
Goldstein HB: See Sachdev VP; Savitz SI
Gordon GV: Arthritis of the lumbar spine, 58(2):109–114
Gorlin R: see Barbasch A
Grand Rounds: Bone marrow failure disorders, B Alter, 58(6):521–534; Doctor's dilemmas: bioethical challenges, A Barbasch, moderator, R Gorlin, D Callahan, B Baumrin, discussants, 58(5):443–451; Interfaces between seizures and affective disorders: evolution and longitudinal course, 58(4):310–323
Graph, of illness course: 58(4):310–323
Green MS: Changing controversy over surgical resection margins for stage I cutaneous melanoma, 58(4):341–346
Grichnik KP, Ferrante FM: Difference between acute and chronic pain, 58(3):217–220

Haberkern CM, Tyler DC, Krane EJ: Postoperative pain management in children, 58(3):247–256
Halperin JL: Prevention of deep venous thrombosis and pulmonary embolism in surgical patients, 58(1):28–36
Hamill RJ: see Rowlingson JC
Hansell DA: AIDS and surrogate decision making, 58(5):375–378
Harrison MJ, Sundaresan N: Spinal instrumentation for degenerative disease of the lumbar spine, 58(2):169–176
Harvey P: see Kramer TA
Hematologic status, preoperative evaluation, JH Rand, C Calandri, 58(1):41–47
Hematology, abstracts, 58(4):358–359
Herniation, lumbar disk: 58(2):129–132, 133–138
Herpes: 58(3):257–266
Hindi AI: see Sachdev VP
History of drug use and drug control, DM Fox, 58(5):403–405
History of pain concepts and pain therapies, JJ Bonica, 58(3):191–202
Holzman IR: Infants and children, surrogacy for, 58(5):382–383
Hypertension: abstracts, 58(4):359

Immunology, clinical: abstracts, 58(4):350–351
Incompetent adults: surrogacy for, 58(5):388–392
Industrial medicine: airway diseases, 58(6):483–493
Infants and children: surrogacy for, IR Holzman, 58(5):382–383; 58(5):393–397; see Children
Infectious diseases: abstracts, 58(4):359–360
In omnia paratus—ready for all things, introduction, Art and science of consultation in medicine, PL Gilbert, 58(1):2
Instrumentation, lumbar spine: 58(2):169–176
Insulin therapy: 58(4):305–309
Intensive care patients: and surrogacy, 58(5):384
Intensive insulin therapy: euglycemia with a price tag, SH Roman, 58(4):305–309
Interfaces between seizures and affective disorders: the uses of visually mapping the evolution and longitudinal course of an illness, RM Post, RL Findling, RS Kahn, 58(4):310–323
Internal medicine: abstracts, 1991, 58(4):347–366; consultation: 58(1):1–94
Internist as preoperative consultant: risk assessment and management, PL Gilbert, 58(1):3–8
Intravenous drug users: and medical ethics, 58(5):403–442
Introductions: Diabetes mellitus: strategies for the 1990s, SH Roman, 58(4):273; In omnia paratus—ready for all things, PL Gilbert, 58(1):2; Issues in medical ethics, DA Moros, R Rhodes, 58(5):367–368; The asthma room, AS Teirstein, 58(6):463–465; Toward the conquest of pain, AP Reed, 58(3):189–190
Issues in medical ethics: drug use; surrogate decision making, DA Moros, R Rhodes, eds, 58(5):367–461

Jacobson L: see Anvaripour PL
Joseph SC: Drug policy in New York City, 58(5):421–422
*Journal:* see *Mount Sinai Journal of Medicine*

Kahn RS: see Post RM
Kambin P: Arthroscopic microdiskectomy, 58(2):159–164
Kaplan AS, Buchbinder D, McCormick S: Temporomandibular disorders and facial pain, 58(3):229–234

Kasachkoff T: Paternalism and drug abuse, 58(5):412–415
Kattan M: see Lowenthal D
Kaye Hon. JS: Staking out the law on surrogacy, 58(5):369–374
Kidney diseases: see Renal disease
King SA: Psychological aspects of pain, 58(3):203–207
Klein JT, Rosen MJ: Pulmonary status, preoperative evaluation, 58(1):37–40
Kramer TA, Merriam A, Harvey P, Lisoskie MD: Acquired brain lesions and psychiatric illness: possible role of dopaminergic systems, 58(4):324–327
Krane EJ: see Haberkern CM
Kraus H, Fischer AA: Diagnosis and treatment of myofascial pain, 58(3):235–239
Kreitzer JM, Reuben SS, Reed AP: Update on postoperative pain management, 58(3):240–246
Kupfer S, editorials: Art and science of consultation in medicine, 58(1):1; The future of medicine, 58(6):535

Laminectomy: 58(2):133–138
Law and surrogate decision making, 58(5):369–381
Lee LV: see Savitz SI
Leibowitz AB: Patients in intensive care, surrogacy for, 58(5):384
Leibowitz AB, Silverstein JH: Anesthesia: what the internist needs to know, 58(1):9–18
Linekin PL: Teaching diabetes self-management skills: a high technology approach, 58(4):293–298
Lisoskie MD: see Kramer TA
Liver diseases: 58(1):75–80; abstracts, 58(4):360–361
Low back pain in children, BR Rosenblum, AS Rothman, 58(2):115–120
Lowenthal D, Kattan M: Asthma in children and its management: an update, 58(6):500–506
Lumbar disk disease: controversies in neurosurgery, MH Savitz, ed, 58(2):95–187
Lumbar disk herniation: 58(2):129–132, 133–138
Lumbar diskectomy using the operating microscope, VP Sachdev, 58(2):147–150
Lumbar facet joint, and synovial cysts: 58(2):125–128
Lumbar spine: arthritis, 58(2):109–114; instrumentation, 58(2):169–176
Lumbar stenosis, LI Malis, 58(2):121–124

McCormick S: see Kaplan HS
Malis LI: Lumbar stenosis, 58(2):121–124
Management, conservative, low back pain: 58(2):97–100, 101–108; 58(3):221–228
Management of low back pain by analgesics and adjuvant drugs, JR DiPalma, GJ DiGregorio, 58(2):101–108
Mann PS: Drugs and the social body, 58(5):427–432
Maranda M: see Frank B
Marel R: see Frank B
Marom ZM: ed, Asthma: new concepts, 58(6):463–520; Mast cells and asthma: possible mechanisms and therapeutic indications, 58(6):472–482
Mast cells and asthma: possible mechanisms and therapeutic indications, ZM Marom, 58(6):472–482
Medical education, nurses in: 58(1):91–94
Medical ethics, 58(5):367–461; bioethical challenges, 58(5):443–451; and drug use, 58(5):403–442; personal ethics, 58(5):452–454; social ethics, 58(5):455–461; and surrogacy, 58(5):369–402
Medical ethics: a broader definition, SM Wolfe, 58(5):455–461
Meier DE: Elderly patients, surrogacy for, 58(5):385–387
Melanoma, cutaneous, surgical resection margins: 58(4):341–346
Merriam A: see Kramer TA
Microdiskectomy: 58(2):139–146, 165–168
Microdiskectomy—myth, mania, or milestone? An 18-year surgical adventure, RW Williams, 58(2):139–146
Millman RB: Pharmacology of the drugs of abuse and the development of public policy, 58(5):416–421
Minilaminotomy as an alternative to laminectomy or microdiskectomy: ten years' experience, MH Savitz, 58(2):165–168
Molecular medicine: abstracts, 58(4):362
Moral perspectives on surrogate decision making, 58(5):388–397
Moreno JD: Foster parents as surrogates for infants and young children, 58(5):393–397
Moros DA, moderator: Drug use: social and scientific background, panel discussion, 58(5):437–440
Moros DA, Rhodes R: eds, Issues in medical ethics: drug use; surrogate decision making, 58(5):367–461; Introduction: Issues in medical ethics, 58(5):367–368; moderators: The ethics of surrogate decision making, panel discussion, 58(5):398–402
*Mount Sinai Journal of Medicine:* awards, 1990, 58(3):iii; editorials, 58(1):1; 58(6):535; death of David A. Dreiling, editor-in-chief emeritus, 58(6): inside back cover; reviewers, list, 1991, 58(6):541
Mount Sinai School of Medicine: awards and prizes 1990/91, 58(3):ii; Dept of Medicine, 1991, abstracts, 58(4):347–365; Faculty awards and prizes, 58(6):542
Murtaugh JB: Drug use and public policy, 58(5):423–426

Neibart E: Antimicrobial therapy, 58(1):19–27
Neoplastic diseases: abstracts, 58(4):362–363; see Oncology
Neuralgia: 58(3):257–266
Neuroradiologic evaluation, lumbar disk herniation: 58(2):129–132
Neurosurgery: 58(2):95–187
New York State: epidemiology of cocaine use, 58(5):406–411; surrogate decisions on sustaining life, 58(5):379–381
Nociceptive transmissions within the spinal cord, LS Sorkin, 58(3):208–216
Nonnarcotic analgesics and tricyclic antidepressants for the treatment of chronic nonmalignant pain, DM Richlin, 58(3):221–228
Nurses in medical education: 58(1):91–94

Occupational airway diseases, EN Schachter, 58(6):483–493
Oncology, medical, abstracts, 58(4):361; see Neoplastic diseases
Onik G: Automated percutaneous lumbar diskectomy, 58(2):151–158
Organization of the multidisciplinary pain center, JC Rowlingson, RJ Hamill, 58(3):267–272

Pain, 58(3):189–172; acute vs. chronic, 58(3):217–220; concepts, history, 58(3):191–202; facial, 58(3):229–234; low back, 58(2):97–100; multidisciplinary treatment center, 58(3):267–272; myofascial, 58(3):235–239; postoperative management, 58(3):240–246; postoperative management in children, 58(3):247–256; psychology: 58(3):203–207; therapies, history, 58(3):191–202; transmission of, 58(3):208–216
Palmer C: Special issues in the management of the elderly patient with diabetes, 58(4):287–292
Panel discussions: Drug use: public policy, B Baumrin, moderator, 58(5):441–442; Drug use: social and scientific background, DA Moros, moderator, 58(5):437–440; ethics of surrogate decision making, DA Moros, R Rhodes, moderators, 58(5):398–402
Paternalism and drug abuse, T Kasachkoff, 58(5):412–415
Patients in intensive care, surrogacy for, AB Leibowitz, 58(5):384
Patient with cardiac disease, preoperative evaluation, HJ Sklaroff, 58(1):48–57
Patient with chronic renal disease, preoperative evaluation, PL Gilbert, R Stein, 58(1):69–74
Patient with endocrine disease, preoperative evaluation, PL Gilbert, 58(1):58–68
Patient with liver disease, preoperative evaluation, M Conn, 58(1):75–80

Pellegrino ED: Personal ethics of the physician: curing medicine from within, 58(5):452–454
Percutaneous radiofrequency rhizotomy of the lumbar facets: ten years' experience, MH Savitz, 58(2):177–178
Perioperative glucose control: does it really matter? A Stagnaro-Green, 58(4):299–304
Personal ethics of the physician: curing medicine from within, ED Pellegrino, 58(5):452–454
Pharmacology of the drugs of abuse and the development of public policy, RB Millman, 58(5):416–421
Physician as a servant of public policy, B Stimmel, 58(5):433–436
Physicians' experiences in surrogate decision making, 58(5):382–387
Physician-nurse collegiality in the medical school curriculum: exploratory workshop and student questionnaire, PL Anvaripour, L Jacobson, J Schweiger, GK Weissman, 58(1):91–94
Portenoy RK: see Galer BS
Postoperative pain management in children, CM Haberkern, DC Tyler, EJ Krane, 58(3):247–256
Post RM, Findling RL. Kahn RS: Interfaces between seizures and affective disorders: the uses of visually mapping the evolution and longitudinal course of an illness, 58(4):310–323
Prediction and prevention of type I diabetes, G Eisenbarth, 58(4):274–279
Pregnant patient: need for surgery unrelated to pregnancy, SH Cherry, 58(1):81–84
Preoperative and postoperative neuroradiologic evaluation of lumbar disk herniation, R Tash, RC Strauss, MH Savitz, 58(2):129–132
Preoperative evaluation: 58(1):19–90
Prevention of deep venous thrombosis and pulmonary embolism in surgical patients, JL Halperin, 58(1):28–36
Prevention: of diabetes, 58(4):274–279; of geriatric disorders, 58(4):328–340; preoperative: 58(1):19–36
Psychiatric disorders: 58(1):85–90; and dopaminergic systems, 58(4):324–327; visual map of illness course, 58(4):310–323
Psychological aspects of pain, SA King, 58(3):203–207
Public policy on drug use, 58(5):421–436; 437–451
Pulmonary diseases, abstracts, 58(4):363–364; see Asthma
Pulmonary function testing in bronchial asthma: standard and emerging techniques, LK Brown, 58(6):507–520
Pulmonary status, preoperative evaluation, JT Klein, MJ Rosen, 58(1):37–40.

Radiology: 58(2):129–132
Rand JH, Calandri C: Hematologic status, preoperative evaluation, 58(1):41–47
Reed AP: ed: Toward the conquest of pain, 58(3):189–272; Introduction: Toward the conquest of pain, 58(3):189–190; see Kreitzer JM
Renal diseases: 58(1):69–74; abstracts, 58(4):364–365
Reuben SS: see Kreitzer JM
Reviewers, *Journal,* 1991, list, 58(6):541
Rheumatologic diseases, abstracts, 58(4):365
Rhizotomy: 58(2):177–178
Rhodes R: see Moros DA
Richlin DM: Nonnarcotic analgesics and tricyclic antidepressants for the treatment of chronic nonmalignant pain, 58(3):221–228
Richman lecture: Medical ethics, a broader definition, SM Wolfe, 58(5):455–461
Right to die: 58(5):379–381
Risk assessment, 58(1):3–8
Risk of wound infection in lumbar disk surgery, SI Savitz, LV Lee, HB Goldstein, 58(2):179–182
Role of atopy in asthma, K Sperber, 58(6):494–499
Role of laminectomy in lumbar disk herniation, CA Fager, 58(2):133–138
Roman SH: ed, Diabetes mellitus: strategies for the 1990s, 58(4):273–366; Introduction, Diabetes mellitus: strategies for the 1990s, 58(4):273; Approach to the heterogeneity of diabetes in adults, 58(4):280–286; Intensive insulin therapy: euglycemia with a price tag, 58(4):305–309
Rosenblum BR, Rothman AS: Low back pain in children, 58(2):115–120
Rosen MJ: see Klein JT
Rothman AS: see Rosenblum BR
Rouse F: Surrogate decisions in New York State on sustaining life, 58(5):379–381
Rowlingson JC, Hamill RJ: Organization of the multidisciplinary pain center, 58(3):267–272

Sachdev VP: Lumbar diskectomy using the operating microscope, 58(2):147–150
Sachdev VP, Savitz MH, Hindi AI, Goldstein HB: Synovial cysts of the lumbar facet joint, 58(2):125–128
Savitz MH: ed, Lumbar disk disease: controversies in neurosurgery, 58(2):95–187; editorial: Lumbar disk disease: controversies in neurosurgery, 58(2):95–96; Minilaminotomy as an alternative to laminectomy or microdiskectomy: ten years' experience, 58(2):165–168; Percutaneous radiofrequency rhizotomy of the lumbar facets: ten years' experience, 58(2):177–178; see Sachdev VP; Tash R
Savitz SI, Lee LV, Goldstein HB: Risk of wound infection in lumbar disk surgery, 58(2):179–182
Schachter EN: Occupational airway diseases, 58(6):483–493
Schanzer A: An adolescent aspiring surgeon's reflections on *Mortal Lessons: Notes on the Art of Surgery* by Richard Selzer, with comment by Michael W. DiGennaro, 58(6):536–540
Schmeidler J: see Frank B
Schweiger J: see Anvaripour PL
Seizures, and affective disorders: 58(4):310–323
Self-management, in diabetes: 58(4):293–298
Short-term and long-term possibilities for interventive gerontology, GM Fahy, 58(4):328–340
Silverstein JH: see Leibowitz AB
Sklaroff HJ: Patient with cardiac disease, preoperative evaluation, 58(1):48–57
Sorkin LS: Nociceptive transmissions within the spinal cord, 58(3):208–216
Special issues in the management of the elderly patient with diabetes, C Palmer, 58(4):287–292
Sperber K: Role of atopy in asthma, 58(6):494–499
Spinal cord damage: abstracts, 58(4):365
Spinal cord, nociceptive transmission in: 58(3):208–216
Spinal instrumentation for degenerative disease of the lumbar spine, MJ Harrison, N Sundaresan, 58(2):169–176
Stagnaro-Green A: Perioperative glucose control: dose it really matter? 58(4):299–304
Staking out the law on surrogacy, Hon. JS Kaye, 58(5):369–374
Stein R: see Gilbert PL
Stenosis, lumbar: 58(2):121–124
Stimmel B: Physician as a servant of public policy, 58(5):433–436
Strauss RC: see Tash R
Sundaresan N: see Harrison MJ
Surgery: and glucose control, 58(4):299–304; lumbar spine, 58(2):133–177, 183–187; lumbar spine, failure, 58(2):183–187; neurologic, 58(2):95–187; pain management, postoperative, 58(3):240–246; pregnancy and, 58(1):81–84; preoperative evaluation, 58(1):1–94; resection margins for melanoma, 58(4):341–346; wound infection during, 58(2):179–182
Surrogate decision making, 58(5):369–402
Surrogate decision making for incompetent adults: an ethical framework, DW Brock, 58(5):388–392
Surrogate decisions in New York State on sustaining life, F Rouse, 58(5):379–381
Synovial cysts of the lumbar facet joint, VP Sachdev, MH Savitz, AI Hindi, HB Goldstein, 58(2):125–128

T
Tash R, Strauss RC, Savitz MH: Preoperative and postoperative neuroradiologic evaluation of lumbar disk herniation, 58(2):129–132
Teaching diabetes self-management skills: a high-technology approach, PL Linekin, 58(4):293–298
Teirstein AS: Differential diagnosis of asthma, 58(6):466–471; Introduction, Asthma: new concepts, The asthma room, 58(6):463–465
Temporomandibular disorders and facial pain, AS Kaplan, D Buchbinder, S McCormick, 58(3):229–234
Therapy: antimicrobial, 58(1):19–27; for pain, 58(3):189–272
Thrombosis: 58(1):28–36
Toward the conquest of pain, AP Reed, ed, 58(3):189–272
Tyler DC: see Haberkern CM

U
Update on postoperative pain management, JM Kreitzer, SS Reuben, AP Reed, 58(3):240–246

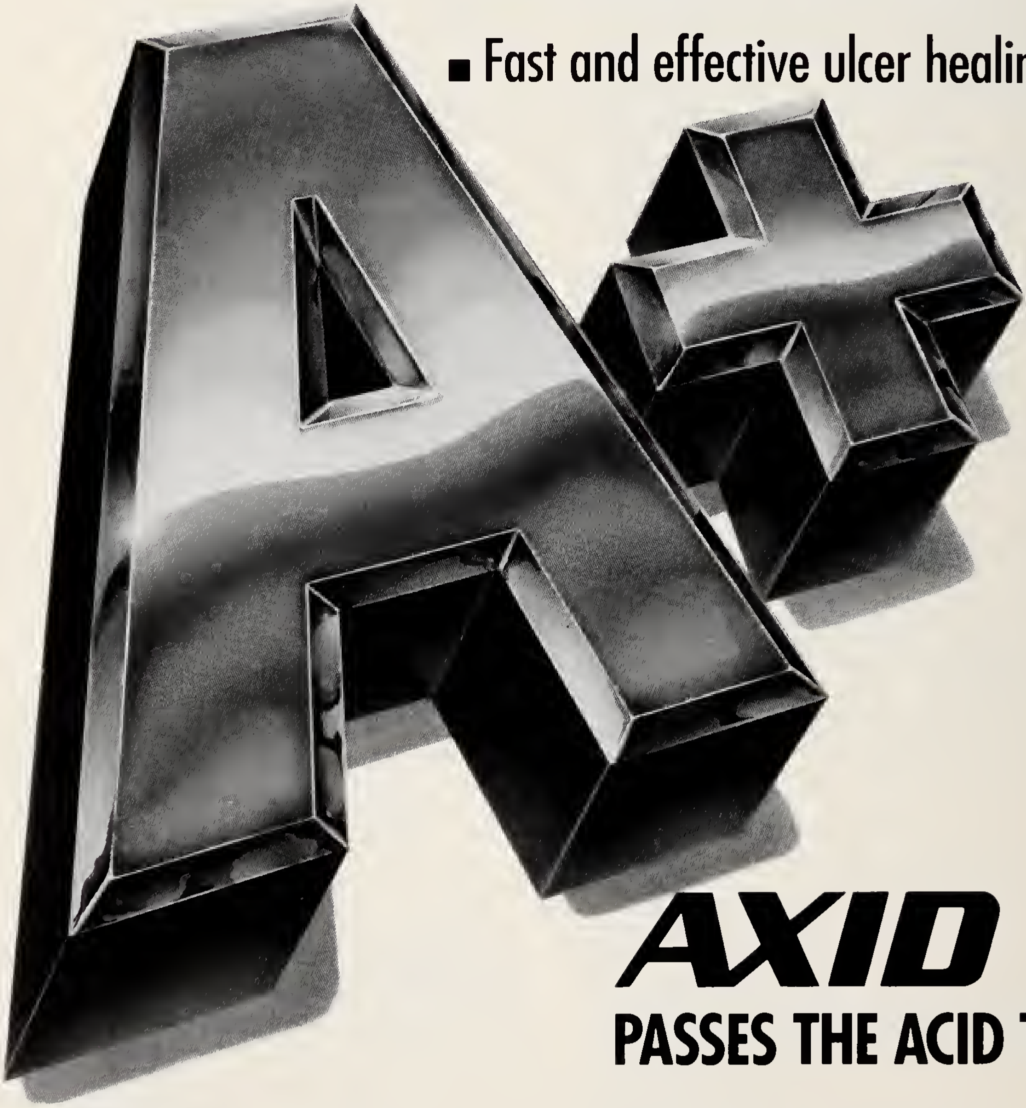
W
Weissman GK: see Anvaripour PL
Williams RW: Microdiskectomy—myth, mania, or milestone? An 18-year surgical adventure, 58(2):139–146
Wolfe SM: Medical ethics: a broader definition, 58(5):455–461
Women patients: surgery during pregnancy, 58(1):81–84
Wound infection in surgery: 58(2):179–182

**For excellent response in the treatment of duodenal ulcers...**

**AXID® nizatidine**

**has the right answers**

- Rapid epigastric pain relief[1,2]*
- Fast and effective ulcer healing[2,3,4]



**AXID**

**PASSES THE ACID TEST**

*Most patients experience pain relief with the first dose.

*See adjacent page for references and brief summary of prescribing information.*

NZ-2943-B-149347

© 1991, ELI LILLY AND COMPANY

**AXID**® (nizatidine capsules)

**Brief Summary. Consult the package insert for complete prescribing information.**
**Indications and Usage** 1 *Active duodenal ulcer*—for up to 8 weeks of treatment Most patients heal within 4 weeks.

2 *Maintenance therapy*—for healed duodenal ulcer patients at a reduced dosage of 150 mg h.s The consequences of therapy with Axid for longer than 1 year are not known

**Contraindications**: Known hypersensitivity to the drug Because cross sensitivity in this class of compounds has been observed, $H_2$-receptor antagonists, including Axid, should not be administered to patients with a history of hypersensitivity to other $H_2$-receptor antagonists

**Precautions**. *General*—1 Symptomatic response to nizatidine therapy does not preclude the presence of gastric malignancy

2 Dosage should be reduced in patients with moderate to severe renal insufficiency

3 In patients with normal renal function and uncomplicated hepatic dysfunction, the disposition of nizatidine is similar to that in normal subjects.

*Laboratory Tests*—False-positive tests for urobilinogen with Multistix® may occur during therapy

*Drug Interactions*—No interactions have been observed with theophylline, chlordiazepoxide, lorazepam, lidocaine, phenytoin, and warfarin Axid does not inhibit the cytochrome P-450 enzyme system, therefore, drug interactions mediated by inhibition of hepatic metabolism are not expected to occur In patients given very high doses (3,900 mg) of aspirin daily, increased serum salicylate levels were seen when nizatidine, 150 mg b.i.d., was administered concurrently.

*Carcinogenesis, Mutagenesis, Impairment of Fertility*—A 2-year oral carcinogenicity study in rats with doses as high as 500 mg/kg/day (about 80 times the recommended daily therapeutic dose) showed no evidence of a carcinogenic effect There was a dose-related increase in the density of enterochromaffin-like (ECL) cells in the gastric oxyntic mucosa In a 2-year study in mice, there was no evidence of a carcinogenic effect in male mice, although hyperplastic nodules of the liver were increased in the high-dose males as compared with placebo Female mice given the high dose of Axid (2,000 mg/kg/day, about 330 times the human dose) showed marginally statistically significant increases in hepatic carcinoma and hepatic nodular hyperplasia with no numerical increase seen in any of the other dose groups The rate of hepatic carcinoma in the high-dose animals was within the historical control limits seen for the strain of mice used The female mice were given a dose larger than the maximum tolerated dose, as indicated by excessive (30%) weight decrement as compared with concurrent controls and evidence of mild liver injury (transaminase elevations) The occurrence of a marginal finding at high dose only in animals given an excessive and somewhat hepatotoxic dose, with no evidence of a carcinogenic effect in rats, male mice, and female mice (given up to 360 mg/kg/day, about 60 times the human dose), and a negative mutagenicity battery are not considered evidence of a carcinogenic potential for Axid

Axid was not mutagenic in a battery of tests performed to evaluate its potential genetic toxicity, including bacterial mutation tests, unscheduled DNA synthesis, sister chromatid exchange, mouse lymphoma assay, chromosome aberration tests, and a micronucleus test.

In a 2-generation, perinatal and postnatal fertility study in rats, doses of nizatidine up to 650 mg/kg/day produced no adverse effects on the reproductive performance of parental animals or their progeny

*Pregnancy—Teratogenic Effects—Pregnancy Category C*—Oral reproduction studies in rats at doses up to 300 times the human dose and in Dutch Belted rabbits at doses up to 55 times the human dose revealed no evidence of impaired fertility or teratogenic effect; but, at a dose equivalent to 300 times the human dose, treated rabbits had abortions, decreased number of live fetuses, and depressed fetal weights On intravenous administration to pregnant New Zealand White rabbits, nizatidine at 20 mg/kg produced cardiac enlargement, coarctation of the aortic arch, and cutaneous edema in 1 fetus, and at 50 mg/kg, it produced ventricular anomaly, distended abdomen, spina bifida, hydrocephaly, and enlarged heart in 1 fetus. There are, however, no adequate and well-controlled studies in pregnant women It is also not known whether nizatidine can cause fetal harm when administered to a pregnant woman or can affect reproduction capacity Nizatidine should be used during pregnancy only if the potential benefit justifies the potential risk to the fetus

*Nursing Mothers*—Studies in lactating women have shown that 0 1% of an oral dose is secreted in human milk in proportion to plasma concentrations Because of growth depression in pups reared by treated lactating rats, a decision should be made whether to discontinue nursing or the drug, taking into account the importance of the drug to the mother

*Pediatric Use*—Safety and effectiveness in children have not been established

*Use in Elderly Patients*—Healing rates in elderly patients were similar to those in younger age groups as were the rates of adverse events and laboratory test abnormalities Age alone may not be an important factor in the disposition of nizatidine Elderly patients may have reduced renal function

**Adverse Reactions**: Clinical trials of varying durations included almost 5,000 patients Among the more common adverse events in domestic placebo-controlled trials of over 1,900 nizatidine patients and over 1,300 on placebo, sweating (1% vs 0.2%), urticaria (0 5% vs <0 01%), and somnolence (2 4% vs 1 3%) were significantly more common with nizatidine It was not possible to determine whether a variety of less common events were due to the drug

*Hepatic*—Hepatocellular injury (elevated liver enzyme tests or alkaline phosphatase) possibly or probably related to nizatidine occurred in some patients In some cases, there was marked elevation (>500 IU/L) in SGOT or SGPT and, in a single instance, SGPT was >2,000 IU/L The incidence of elevated liver enzymes overall and elevations of up to 3 times the upper limit of normal, however, did not significantly differ from that in placebo patients All abnormalities were reversible after discontinuation of Axid Since market introduction, hepatitis and jaundice have been reported Rare cases of cholestatic or mixed hepatocellular and cholestatic injury with jaundice have been reported with reversal of the abnormalities after discontinuation of Axid

*Cardiovascular*—In clinical pharmacology studies, short episodes of asymptomatic ventricular tachycardia occurred in 2 individuals administered Axid and in 3 untreated subjects

*CNS*—Rare cases of reversible mental confusion have been reported

*Endocrine*—Clinical pharmacology studies and controlled clinical trials showed no evidence of antiandrogenic activity due to nizatidine Impotence and decreased libido were reported with equal frequency by patients on nizatidine and those on placebo Gynecomastia has been reported rarely

*Hematologic*—Fatal thrombocytopenia was reported in a patient treated with nizatidine and another $H_2$-receptor antagonist This patient had previously experienced thrombocytopenia while taking other drugs Rare cases of thrombocytopenic purpura have been reported

*Integumental*—Sweating and urticaria were reported significantly more frequently in nizatidine- than in placebo-treated patients Rash and exfoliative dermatitis were also reported

*Hypersensitivity*—As with other $H_2$-receptor antagonists, rare cases of anaphylaxis following nizatidine administration have been reported Rare episodes of hypersensitivity reactions (eg, bronchospasm, laryngeal edema, rash, and eosinophilia) have been reported

*Other*—Hyperuricemia unassociated with gout or nephrolithiasis was reported Eosinophilia, fever, and nausea related to nizatidine have been reported

**Overdosage** Overdoses of Axid have been reported rarely If overdosage occurs, activated charcoal, emesis, or lavage should be considered along with clinical monitoring and supportive therapy Renal dialysis does not substantially increase clearance of nizatidine due to its large volume of distribution

PV 2091 AMP
[091190]

**References**
1 Data on file, Lilly Research Laboratories
2 *Scand J Gastroenterol* 1987,22(suppl 136) 61-70
3 *Scand J Gastroenterol* 1987,22(suppl 136) 47-55
4 *Am J Gastroenterol* 1989,84 769-774

NZ-2943 B 149347

*Additional information available to the profession on request*

Lilly

**Eli Lilly and Company**
Indianapolis, Indiana
46285

The editorial staff of
The Mount Sinai Journal of Medicine
notes with sadness
the death of David A. Dreiling,
editor-in-chief from 1974 to 1990
and editor emeritus
since the beginning of this year,
on September 24, 1991.

Comments by his professional
colleagues and friends
will be published in a future issue.

# THE MOUNT SINAI JOURNAL OF MEDICINE

LEVY LIBRARY, MT. SINAI MEDICAL CENTER
3 4805 0327993 8

## Forthcoming: Grand Rounds

**Graves Disease**
Terry Davies, MD
Mount Sinai School of Medicine

**Is There A Future for Physicians in America?**
Uwe E. Reinhardt, MD
Princeton University

**Osteoporosis: On the Verge of Rational Effective Therapy?**
John Paul Bilezikian, MD
Columbia University College of Physicians and Surgeons

**Obscure Gastrointestinal Bleeding**
Blair S. Lewis, MD
Mount Sinai School of Medicine

**Adrenergic Receptors in Hypertension and Heart Failure**
Clive Rosendorff, MD
University of the Witwatersrand Medical School

**Why Kidneys Fail**
Lance Dworkin, MD
New York University School of Medicine

## Theme Issues

### MARCH 1992

**Hormones, Metabolism, and Society: In Honor of Rosalind Yalow**
*Editor:* Eugene Straus

### SEPTEMBER 1992

**Human Cancer: From Precurable to Curable**
*Editors:* Max Sung and Takao Ohnuma

### NOVEMBER 1992

**Community Medicine: In Honor of Kurt Deuschle**
*Editor:* Alan Silver

### MARCH 1993

**Controversies in Neurosurgery: Cervical Spine Disorders**
*Editors:* Martin Savitz and Kalmon Post

## Available now:

**Toward the Conquest of Pain**
Allan P. Reed, editor
volume 58, no. 3, May 1991
84 pages **$15**

**Lumbar Disk Disease: Controversies in Neurosurgery**
Leonard I. Malis and Martin H. Savitz, editors
volume 58, no. 2, March 1991
96 pages **$15**

**The Art and Science of Consultation in Medicine**
Paul L. Gilbert, editor
volume 58, no. 1, January 1991
96 pages **$15**

*Subscriptions for 1991: $60 individuals, U.S. and Canada, and $65 all libraries; $65 individuals outside U.S. and Canada.*

*To order subscriptions or copies of back issues ($15 each), please send check, payable to The Mount Sinai Journal of Medicine, to the Journal at Box 1094, One Gustave L. Levy Place, New York, NY 10029-6574; counter sales at Suite 4D, 19 E. 98th Street, New York, NY (phone: 212 241-6108; FAX: 212 722-6386).*

MOUNT SINAI
GUSTAVE L.
AND
JANET W.
LEVY
LIBRARY

LEVY LIBRARY, MT. SINAI MEDICAL CENTER
3 4805 03340835 4